Abstract of Paper from ASONAM 2012, FOSINT-SI 2012, HI-BI-BI 2012 and all workshops

Large Social Networks Can Be Targeted for Viral Marketing with Small Seed Sets

Paulo Shakarian, Damon Paulo

In a "tipping" model, each node in a social network, representing an individual, adopts a behavior if a certain number of his incoming neighbors previously held that property. A key problem for viral marketers is to determine an initial "seed" set in a network such that if given a property then the entire network adopts the behavior. Here we introduce a method for quickly finding seed sets that scales to very large networks. Our approach finds a set of nodes that guarantees spreading to the entire network under the tipping model. After experimentally evaluating 31 real-world networks, we found that our approach often finds such sets that are several orders of magnitude smaller than the population size. Our approach also scales well - on a Friendster social network consisting of 5.6 million nodes and 28 million edges we found a seed sets in under 3.6 hours. We also find that highly clustered local neighborhoods and dense network-wide community structure together suppress the ability of a trend to spread under the tipping model.

The Trilemma of Network Analysis

Isadora Dorn, Andreas Lindenblatt, Katharina A. Zweig

The recent interest in network analysis is caused by the unprecedented accessibility to large datasets: there are huge, publicly available databases on protein-protein-interactions, air transportation, and street maps which easily lend themselves to a network representation. Once a network is created, all types of path-based network analytic measures can be easily applied: typical examples are centrality measures, but also some clustering algorithms and robustness analysis rely on path-based measures. Borgatti has claimed that centrality measures basically simulate dissemination processes of goods which use a certain subset of paths on the given network [1]; they can thus only be used to describe processes which rely on the same type of good and the same subset of paths. Later, Butts pointed out that the results of a chosen network analytic method strongly vary with modeling decisions taken when turning raw data into networks [2]. In this article we combine these two insights to the trilemma of network analysis which states that the network process of interest, the network representation, and the network analytic method cannot be chosen independently. We discuss on two real-world examples in the realm of air transportation networks how to choose a distance based measure with respect to the context of the data, re-computing similar analyses by Guimer´a et al. [3] and Dall'Asta et al. [4]. In both cases, the path-based measures matching the network process of interest change the interpretation of the previous findings, which shows the potential in regarding the trilemma of network analysis.

Link and Node Analysis of Gender Based Collaborations in Turkish Social Sciences

Bülent Özel

This paper examines impact of gender both on publication productivity and on patterns of scientific collaborations in social sciences in Turkey. Bibliographic data on local publications in Turkey is used. It consists of 7835 papers written by 6738 scientists. Findings suggest that (1) there are gender differences at publication productivity, participation, presence and contribution, that (2) there are significantly different tendencies at keeping established co-authorship ties for inter-gender and intra-gender pairs, that (3) there are significant gender differences at positions of individuals in the network structure. It is seen that while female scientists are seen to be embedded in cliques more often than males, males are more active at bridging different components in the network. This study exemplifies an integrated approach to better examine role of gender in scientific collaborations.

Semantic Expansion of Tweet Contents for Enhanced Event Detection in Twitter

Ozer Ozdikis, Pinar Senkul, Halit Oguztuzun

This paper aims to enhance event detection methods in a micro-blogging platform, namely Twitter. The enhancement technique we propose is based on lexico-semantic expansion of tweet contents while applying document similarity and clustering algorithms. Considering the length limitations and idiosyncratic spelling in Twitter environment, it is possible to take advantage of word similarities and to enrich texts with similar words. The semantic expansion technique we implement is based on syntagmatic and paradigmatic relationships between words, extracted from their co-occurrence statistics. As our technique does not depend on an existing ontology or a lexicon database such as Word Net, it should be applicable for any language. The proposed technique is applied on a tweet set collected for three days from the users in Turkey. The results indicate earlier detection of events and improvements in accuracy.

Abstract of Paper from ASONAM 2012, FOSINT-SI 2012, HI-BI-BI 2012 and all workshops

Social-Based Conceptual Links: Conceptual Analysis Applied to Social Networks

Erick Stattner, Martine Collard

In this work, we propose a novel approach for the discovery of frequent patterns in a social network on the basis of both vertex attributes and link frequency. With an analogy to the traditional task of mining frequent item sets, we show that the issue addressed can be formulated in terms of a conceptual analysis that elicits conceptual links. A social-based conceptual link is a synthetic representation of a set of links between groups of vertexes that share similar internal properties. We propose a first algorithm that optimizes the search into the concept lattice of conceptual links and extracts maximal frequent conceptual links. We study the performances of our solution and give experimental results obtained on a sample example. Finally we show that the set of conceptual links extracted provides a conceptual view of the social network.

Evolutionary Community Detection for Observing Covert Political Elite Cliques

Jyi-Shane Liu, Ke-Chih Ning, Wan-Chun Chuang

Among many real world applications of social network analysis, political interaction and executive succession show some unique characteristics of dynamic community evolution and raise interesting research challenges. Interactions of political power among community members are mostly subtle and behind the scene. Visible relations are only nominal and are not readily apparent to key findings. Under such difficult circumstances of information deficiency, the research problem is to uncover the inner relations among some of the network entities and to discover the hidden network structure based on these inner relations. In this research, our objective is to identify the inner circles of government political power and bureaucracy underneath formal work relations and observe how the political elite groups form and change over time. A government official job change network in a time span of over twenty years is built to model synchronous post assignment and job promotion within a time window as entity relations. In each snapshot of network evolution, communities that exhibit strong association of synchronous job change are identified by the edge betweenness decomposition algorithm. Then, an event-based framework is used to characterize community behavior patterns in consecutive changes of network structures. The approach is effectually demonstrated on two scenarios: (1) identifying and tracking the inner circle of a leading political figure, (2) finding succession pool members in government agencies. We further propose two evolutionary community variation indexes to assess political executive succession. Experimental results with actual government personnel data provide evidence that government agency succession can be reasonably measured. This work also has the practical value of providing objective scrutiny on political power transition for the benefit of public interest.

Measuring Topological Robustness of Networks under Sustained Targeted Attacks

Mahendra Piraveenan, Shahadat Uddin, Kon Shing Kenneth Chung

In this paper, we introduce a measure to analyse the structural robustness of complex networks, which is specifically applicable in scenarios of targeted, sustained attacks. The measure is based on the changing size of the largest component as the network goes through disintegration. We argue that the measure can be used to quantify and compare the effectiveness of various attack strategies. Applying this measure, we confirm the result that scale-free networks are comparatively less vulnerable to random attacks and more vulnerable to targeted attacks. Then we analyse the robustness of a range of real world networks, and show that most real world networks are least robust to attacks based on betweenness of nodes. We also show that the robustness of some networks are more sensitive to the attack strategy compared to others, and given the disparity in the computational complexities of calculating various centrality measures, the robustness coefficient introduced can play a key role in choosing the attack and defence strategies for real world networks. While the measure is applicable to all types of complex networks, we clearly demonstrate its relevance to social network analysis.

Analyzing User Retweet Behavior on Twitter

Zhiheng Xu, Qing Yang

This paper provides a deep analysis of user retweet behavior on Twitter. While previous works about analyzing retweet have mainly focused on predicting the retweetability of each tweet, they lacked interpretations at an individual level. In this paper, we perform a general analysis of retweet behavior from the perspective of individual users. Specifically, we train a prediction model to forecast whether a tweet will be retweeted by a given user, leveraging four different types of features: social-based, content-based, tweet-based and author-based features. By performing ¡°leave-one-feature-out¡± comparisons, we identify factors that are strongly associated with user retweet behavior.

Applying SNA for the Characterization of the Spatial Organization of XiaoLin Village

Yun-Shang Chiou, Ying-Ting Huang

Through site visits, interviews, geographical data collection, and periodical update of settlements¡¦ conditions, this study reconstructs the spatial organization of XiaoLin village in southern Taiwan, destroyed by Typhoon Maroko in 2009. The spatial organizations of the dwellings of the original and new settlements of XiaoLin villagers were analyzed by measures from social network analysis and compared against each other to see whether the ¡§consensus¡¨ in spatial organization exist and how it responds to villagers' life style. A sustainable community development seeks to balance its environmental, economical, social and cultural needs. In recent years, governments and NGOs in Taiwan invest huge amounts of resources in post-disaster community reconstruction. Their focuses are mainly on environmental and economical fronts. What has often been overlooked is the preservation of the way of life in the original habitat. By applying SNA measures, this study extracts the characteristics of the spatial organization that facilitate the fabrics of everyday life in a human settlement and proposes how this information can contribute to social and cultural preservation in Taiwan¡¦s post-disaster reconstruction in the future.

Finding a Maximum Clique Using Ant Colony Optimization and Particle Swarm Optimization in Social Networks

Mohammad Soleimani-Pouri, Alireza Rezvanian, Mohammad Reza Meybodi

Interaction between users in online social networks plays a key role in social network analysis. One on important types of social group is full connected relation between some users, which known as clique structure. Therefore finding a maximum clique is essential for some analysis. In this paper, we proposed a new method using ant colony optimization algorithm and particle swarm optimization algorithm. In the proposed method, in order to attain better results, it is improved process of pheromone update by particle swarm optimization. Simulation results on popular standard social network benchmarks in comparison standard ant colony optimization algorithm are shown a relative enhancement of proposed algorithm.

Microblogging in the Enterprise: A Few Comments are in Order

Charalampos Chelmis, Viktor K. Prasanna

Popular social networking sites have revolutionized the way people interact on the Web. Researchers have studied social networks from numerous perspectives, mostly focusing on publicly available social networks and microblogging sites. Enterprises however have recently being adopting and utilizing microblogging services as part of their day to day operations. The goal of this paper is to study the topological properties of a corporate microblogging service, its dynamics and characteristics. Through an extensive analysis of enterprise microblogging data, we provide insights on the structural properties of the extracted network of directed messages sent between users of a corporate microblogging service, as well as the lexical and topical alignment of users. We compare our results to traditional, general purpose, online social networks and discuss the implications of our findings. To the best of our knowledge, this work is the first quantitative study of an enterprise microblogging service, its usage characteristics, and its derived social network based on replies between users.

Stock Market Investment Advice: A Social Network Approach

Negar Koochakzadeh, Keivan Kianmehr, Atieh Sarraf, Reda Alhajj

Making investment decision on various available stocks in the market is a challenging task. Econometric and statistical models, as well as machine learning and data mining techniques, have proposed heuristic based solutions with limited long-range success. In practice, the capabilities and intelligence of financial experts is required to build a managed portfolio of stocks. However, for non-professional investors, it is too complicated to make subjective judgments on available stocks and thus they might be interested to follow an expert's investment decision. For this purpose, it is critical to find an expert with similar investment preferences. In this work, we propose to benefit from the power of Social Network Analysis in this domain. We first build a social network of financial experts based on their publicly available portfolios. This social network is then used for further analysis to recommend an appropriate managed portfolio to non-professional investors based on their behavioral similarities to the expert investors. This approach is evaluated through a case study on real portfolios. The result shows that the proposed portfolio recommendation approach works well in terms of Sharpe ratio as the portfolio performance metric.

Global Similarity in Social Networks with Typed Edges

D.B. Skillicorn, Q. Zheng

Most real-world social network analysis treats edges (relationships) as having different intensities (weights), but the same qualitative properties. We address the problem of modelling edges of qualitatively different types that nevertheless interact with one another. For example, influence flows along friend and colleagues edges differently, but treating the two sets of different kinds of edges as independent graphs surely misses interesting and useful structure. We model the sub graph corresponding to each edge type as a layer, and show how to weight the edges connecting the layers to produce a consistent spectral embedding, including for directed graphs. This embedding can be used to compute social network properties of the combined graph, to predict edges, and to predict edge types. We illustrate with Padgett's dataset of Florentine families in the 15th Century.

Measuring the Importance of Users in a Social Network Based on Email Communication Patterns

Pawel Lubarski, Mikolaj Morzy

Email communication patterns have been long used to derive the underlying social network structure. By looking at who is talking to who and how often, the researchers have disclosed interesting patterns, hinting on social roles and importance of actors in such networks. Email communication analysis has been previously applied to discovering cliques and fraudulent activities (e.g. the Enron email network), to observe information dissemination patterns, and to identify key players in email communication networks. In this paper we are using a large dataset of email communication within a constrained community (i.e. the employees of a single institution) to discover the importance of employees in the underlying network. Contrary to previous attempts, though, we are scrutinizing the delays in answering emails. We base our method on a simple notion of implicit importance: people are more likely to quickly respond to emails sent by people who are being perceived as important. In the paper we propose several methods for building the social network from the email communication data and we introduce various weighting schemes. We aggregate the resulting rankings and compute differences between rankings to observe the stability of our method. We also compare the resulting rankings with an a priori assessment of employees' importance to verify our method. The results of the conducted experiments clearly show the validity and robustness of the proposed method.

Network of Practices: A Case Study of Knowledge Competition of School Technology Coordinators

Fang-Ling Lin, Guey-Fa Chiou

This study investigates the network problems of the knowledge dissemination of ICT-in-education practices using an egocentric survey method to school technology coordinators (STCs). A conceptual framework of knowledge competitions is developed to inspect the significant and effective ties of knowledge dissemination. The results indicate that interactive predictability of effective sizes and significant ties to knowledge competitions can reveal the effects on competitive advantages. Based on this predictability, a person's knowledge-contribution to a network of ICT-integration instruction is classified as ineffective with less positional and informational benefits. These findings help confront the deficiency of the achievements of ICT integration instructions and encourage reconsidering the role of STCs in ICT-in-education practices.

A Multi-Classifier System for Sentiment Analysis and Opinion Mining

Luana Bezerra Batista, Sylvie Ratté

Although successfully employed to reduce error rates of difficult pattern recognition problems, multi-classifier systems (MCS) are not in widespread use in the field of Sentiment Analysis and Opinion Mining. The motivation of using a MCS stems from the fact that different classifiers usually make different errors on different samples. By using just the best classifier, it is possible to loose valuable information contained in the other sub optimal classifiers. In this work, we take advantage of unigrams, big rams and trig rams to design a multi-classifier system for Sentiment Analysis and Opinion Mining. Three different Naive Bayes classifiers are trained – each one with a specific set of features –, and then combined in the ROC space by using the Iterative Boolean Combination (IBC) technique. IBC iteratively combines the ROC curves produced by different classifiers using all Boolean functions, and does not require prior assumption that the classifiers are statistically independent. An experimental study investigates the advantage of using the proposed MCS, over each individual classifier, in classifying Twitter messages as positive or negative. The Stanford University's Twitter database is employed for this task. As real-world application, the proposed MCS is used to identify the sentiment of electors regarding the main candidates for the 2012 United States Presidential Elections. Results indicate that the proposed MCS can provide useful information about people's opinions that are comparable to conventional opinion polls.

On Measurement of Influence in Social Networks

Behnam Hajian, Tony White

One of the issues to be resolved in social recommender systems is the identification of opinion leaders in a network. Finding effective people in societies has been a key question for many groups, e.g., marketers. The research undertaken in this paper focuses on finding important nodes in a network based on their behaviour as well as the structure of the network. This paper views the propagation of information in a social network as a process of infection. The paper proposes an algorithm called the Probability Propagation Method for measuring the probability of infection of all the nodes in a network starting from a given node in the network. Then, assuming independence in activation of nodes in a network, a method is proposed for ranking nodes according to their capabilities in infecting a larger number of nodes in a network. These methods are validated using simulation software in which a non-deterministic model of information diffusion is simulated on several classes of network.

Ranking News Articles Based on Popularity Prediction

Alexandru Tatar, Panayotis Antoniadis, Marcelo Dias de_Amorim, Serge Fdida

News articles are a captivating type of online content that capture a significant amount of Internet users' interest. They are particularly consumed by mobile users and extremely diffused through online social platforms. As a result, there is an increased interest in promptly identifying the articles that will receive a significant amount of user attention. This task falls under the broad scope of content popularity prediction and has direct implications in various contexts such as caching strategies or online advertisement policies. In this paper we address the problem of predicting the popularity of news articles based on user comments. We formulate the prediction task into a ranking problem where the goal is not to infer the precise attention that a content will receive but to accurately rank articles based on their predicted popularity. To this end, we analyze the ranking performance of three prediction models using a dataset of articles covering a four-year period and published by 20minutes.fr, an important French online news platform. Our results indicate that prediction methods improve the ranking performance and we observed that for our dataset a simple linear prediction method outperforms more dedicated prediction methods.

@Phillies Tweeting from Philly? Predicting Twitter User Locations with Spatial Word Usage

Hau-wen Chang, Dongwon Lee, Mohammed Eltaher, Jeongkyu Lee

We study the problem of predicting home locations of Twitter users using contents of their tweet messages. Using three probability models for locations, we compare both the Gaussian Mixture Model (GMM) and the Maximum Likelihood Estimation (MLE). In addition, we propose two novel unsupervised methods based on the notions of Non-Localness and Geometric-Localness to prune noisy data from tweet messages. In the experiments, our unsupervised approach improves the baselines significantly and shows comparable results with the supervised state-of-the-art method. For 5,113 Twitter users in the test set, on average, our approach with only 250 selected local words or less is able to predict their home locations (within 100 miles) with the accuracy of 0.499, or has 509.3 miles of average error distance at best.

Analyzing Stock Market Movements Using Twitter Sentiment Analysis

Tushar Rao, Saket Srivastava

In this paper we investigate the complex relationship between tweet board literature (like bullishness, volume, agreement etc) with the financial market instruments (like volatility, trading volume and stock prices). We have analyzed sentiments for more than 4 million tweets between June 2010 to July 2011 for DJIA, NASDAQ-100 and 13 other big cap technological stocks. Our results show high correlation (up to 0.88 for returns) between stock prices and twitter sentiments. Further, using Grangerâ€™s Causality Analysis, we have validated that the movement of stock prices and indices are greatly affected in the short term by Twitter discussions. Finally, we have implemented Expert Model Mining System (EMMS) to demonstrate that our forecasted returns give a high value of Rsquare (0.952) with low Maximum Absolute Percentage Error (MaxAPE) of 1.76% for Dow Jones Industrial Average (DJIA).

Predicting User-to-content Links in Flickr Groups

Sumit Negi, Santanu Chaudhury

The last few years have seen an exponential increase in the amount of multimedia content that is available online thanks to collaborative-online communities such as Flickr, You Tube etc. As opposed to "pure" social networking services these collaborative-online communities not only allow users to create new social links (e.g. add other users to their friend or contact list) but also allow users to contribute multimedia content and engage in content-driven interactions (called user-to-content interactions). A good example of this can be seen in Flickr, in general and Flickr Group in particular where users can comment on or "like" an image contributed by another user. This paper looks at the task of predicting the formation of such user-to-content links in Flickr Groups. More specifically, "what is the chance that a user will comment/like an image contributed by another user?". Our proposed method for predicting user-to-content links takes into account both community effect and content effect. Our results on real-world Flickr Group data reveals that the proposed method shows good performance for the user-to-content link prediction task.

What's in Twitter: I Know What Parties are Popular and Who You are Supporting Now!

Antoine Boutet, Hyoungshick Kim, Eiko Yoneki

In modern politics, parties and individual candidates must have an online presence and usually have dedicated social media coordinators. In this context, we study the usefulness of analysing Twitter messages to identify both the characteristics of political parties and the political leaning of users. As a case study, we collected the main stream of Twitter related to the 2010 UK General Election during the associated period – gathering around 1,150,000 messages from about 220,000 users. We examined the characteristics of the three main parties in the election and highlighted the main differences between parties. First, Lab our members were the most active and influential during the election while Conservative members were the most organized to promote their activities. Second, the websites and blogs that each political party's members supported are clearly different from those that all the other political parties' members supported. From these observations, we develop a simple and practical classification method which uses the number of Twitter messages referring to a particular political party. The experimental results showed that the proposed classification method achieved about 86% classification accuracy and outperforms other classification methods that require expensive costs for tuning classifier parameters and/or knowledge about network topology.

Analyzing Voting Behavior in Italian Parliament: Group Cohesion and Evolution

Alessia Amelio, Clara Pizzuti

The roll calls of the Italian Parliament in the current legislature is studied by employing multidimensional scaling, hierarchical clustering, and network analysis. In order to detect changes in voting behavior, the roll calls have been divided in seven periods of six months each. All the methods employed pointed out an increasing fragmentation of the political parties endorsing the previous government that culminated in its downfall. By using the concept of modularity at different resolution levels, we identify the community structure of Parliament and its evolution in each of the time periods considered. The analysis performed revealed as a valuable tool in detecting trends and drifts of Parliamentarians. It showed its effectiveness at identifying political parties and at providing insights on the temporal evolution of groups and their cohesiveness, without having at disposal any knowledge about political membership of Representatives.

Comparison of Co-authorship Networks across Scientific Fields Using Motifs

Sarvenaz Choobdar, Pedro Ribeiro, Sylwia Bugla, Fernando Silva

Comparing scientific production across different fields of knowledge is commonly controversial and subject to disagreement. Such comparisons are often based on quantitative indicators, such as papers per researcher, and data normalization is very difficult to accomplish. Different approaches can provide new insight and in this paper we focus on the comparison of different scientific fields based on their research collaboration networks. We use co-authorship networks where nodes are researchers and the edges show the existing co-authorship relations between them. Our comparison methodology is based on network motifs, which are over represented patterns, or sub graphs. We derive motif fingerprints for 22 scientific fields based on 29 different small motifs found in the corresponding co-authorship networks. These fingerprints provide a metric for assessing similarity among scientific fields, and our analysis shows that the discrimination power of the 29 motif types is not identical. We use a co-authorship dataset built from over 15,361 publications inducing a co-authorship network with over 32,842 researchers. Our results also show that we can group different fields according to their fingerprints, supporting the notion that some fields present higher similarity and can be more easily compared.

Link Prediction for Bipartite Social Networks: The Role of Structural Holes

Shuang Xia, BingTian Dai, Ee-Peng Lim, Yong Zhang, Chunxiao Xing

Link prediction is an important problem in social network mining. Traditional neighborhood based methods such as Common neighbors, Jaccard Coefficient and Adamic/Adar are well studied in link prediction. However, the concept of structural holes does not receive significant attention in link prediction. As a preliminary work in studying structural holes, we focus on bipartite social networks, which is a special class of social networks that consists of two distinct roles for the users, and links are between users of different roles. In this study, a few implementations of structural holes are proposed, which are then validated with extended neighborhood based methods on a real dataset derived from IMDb network. The results show that structural holes help in improving accuracies in link prediction.

Where's the Money? The Social Behavior of Investors in Facebook's Small World

Liang Yuxian Eugene, Soe-Tsyr Daphne Yuan

Are investing activities dependent on social relationships? In our research, we apply social network analysis to the field of investing behaviors and discover that investors have a tendency to invest in companies that are socially similar to them. While traditional studies on investing behavior tend to focus on factors like psychology, opinions, investing experience etc, they fail to consider social relationship as an important factor. In this paper we provide general rules of thumb that are useful for companies seeking funding from investor. These rules of thumb are generated by analyzing the social relationships between investors and companies found within the small world of Facebook.

Subgraph Extraction for Trust Inference in Social Networks

Yuan Yao, Hanghang Tong, Feng Xu, Jian Lu

Trust inference is an essential task in many real world applications. Most of the existing inference algorithms suffer from the scalability issue, making themselves computationally costly, or even infeasible, for the graphs with more than thousands of nodes. In addition, the inference result, which is typically an abstract, numerical trustworthiness score, might be difficult for the end-user to interpret. In this paper, we propose sub graph extraction to address these challenges. The core of the proposed method consists of two stages: path selection and component induction. The outputs of both stages can be used as an intermediate step to speed up a variety of existing trust inference algorithms. Our experimental evaluations on real graphs show that the proposed method can accelerate existing trust inference algorithms, while maintaining high accuracy. In addition, the extracted sub graph provides an intuitive way to interpret the resulting trustworthiness score.

Towards Maximising Cross-Community Information Diffusion

Václav Belák, Samantha Lam, Conor Hayes

In recent years, companies have started to utilise online social communities as a means of communicating with and targeting their employees and customers, and such online communities include discussion fora. The conversational dynamics of users in fora can influence their neighbours in the underlying social network. We make use of such influence to target specific communities with information, i.e. post in them, because a post is generally shared with the community and not just with individual users. In short, we study information diffusion across communities and show that we can achieve high community (and user) spread using an efficient targeting strategy. In order to achieve this, we use a set of novel measures for cross-community influence and show that it outperforms other targeting strategies on two different data-sets: the largest Irish online discussion system, Boards.ie, and technical support fora, SAP SCN.

Visual Analysis of Dynamic Networks Using Change Centrality

Paolo Federico, Jürgen Pfeffer, Wolfgang Aigner, Silvia Miksch, Lukas Zenk

The visualization and analysis of dynamic social networks are challenging problems, demanding the simultaneous consideration of relational and temporal aspects. In order to follow the evolution of a network over time, we need to detect not only which nodes and which links change and when these changes occur, but also the impact they have on their neighbourhood and on the overall relational structure. Aiming to enhance the perception of structural changes at both the micro and the macro level, we introduce the change centrality metric. This novel metric, as well as a set of further metrics we derive from it, enable the pair wise comparison of subsequent states of an evolving network in a discrete-time domain. Demonstrating their exploitation to enrich visualizations, we show how these change metrics support the visual analysis of network dynamics.

Tag Ranking by Linear Relational Neighbourhood Propagation

Boris Chidlovskii

We propose a tag recommendation method which can assist users in tagging process by suggesting relevant tags. % or directly expand the set of tags. The method is based on query-based ranking on relational multi-type graphs which capture the annotation relationship between objects and tags, as well as the object similarity and tag correlation. The additional advance consists in extending the linear neighbourhood propagation to the relational graphs with the Laplacian regularization framework. We report evaluation results on a large-scale Flickr data set.

Examining Multi-factor Interactions in Microblogging Based on Log-linear Modeling

Zhilin Luo, Xintao Wu, Wandong Cai, Dong Peng

Microblogging, as a new form of social media, attracts a huge number of users and becomes very popular. In this paper, we consider a fundamental social network issue that illustrates how information flows through a social media network and specify why users have different retweet behaviors. We propose to characterize social ties by using various features such as power ratio, local link structure, location, and gender. Those features can be directly extracted from users' profiles in Microblogging sites. We apply a fitted Log-linear model to describe association patterns among the features and retweet factor. Using the fitted Log-linear model, we explain why users with different profiles and link structures have different retweet behaviors. Our evaluations on Sina Weibo data set show several phenomenons.

A Bayesian Hierarchical Approach for Exploratory Analysis of Communities and Roles in Social Networks

Gianni Costa, Riccardo Ortale

We present a new probabilistic approach to modeling social interactions, that seamlessly integrates community discovery and role assignment for a deeper understanding of connectivity patterns in social networks. The devised approach is an unsupervised learning technique based on a Bayesian hierarchical model of social interactions. This model specifies an intuitive generative process, in which pairs of nodes in a social network are associated with communities as well as roles in the context of the respective communities, before that a directed interaction is possibly established between them. According to the generative semantics of the proposed model, nodes are represented as probability distributions over communities, while communities are represented as probability distributions over roles. Such distributions are unknown parameters of the proposed model, that are estimated from social-network data through approximated posterior inference and parameter estimation. A comparative evaluation over real-world social networks reveals that our approach outperforms state-of-the-art competitors in terms of link prediction.

User Features and Social Networks for Topic Modeling in Online Social Media

Bo Hu, Zhao Song, Martin Ester

In recent years, social media websites, such as Epinions, Twitter, and Google+, have gained in popularity and have become ubiquitous in our daily lives, where rich user-generated texts are propagated through social networks. Topic models, such as Latent Dirichlet Allocation (LDA), have been proposed and shown to be useful for text analysis. The existing topic models focus on traditional document collections, which consist of a relatively small number of long and high-quality documents. However, user-generated texts tend to be shorter and noisier than traditional content. Besides, the social networks have two novel features: context information on nodes, such as user features, and edges, such as relationship, which have not been considered by the existing topic models. In this paper, we pose the problem of finding user topics in large-scale collection of documents from online social networks. We propose a comprehensive Feature based and a Social based Topic model, taking into account the user features and social networks. We demonstrate that our models have better performance than a baseline LDA in the Epinions, Twitter, and Google+ data sets.

Collective Churn Prediction in Social Network

Richard J. Oentaryo, Ee-Peng Lim, David Lo, Feida Zhu, Philips K. Prasetyo

In service-based industries, churn poses a significant threat to the integrity of the user communities and profitability of the service providers. As such, research on churn prediction methods has been actively pursued, involving either intrinsic, user profile factors or extrinsic, social factors. However, existing approaches often address each type of factors separately, thus lacking a comprehensive view of churn behaviors. In this paper, we propose a new churn prediction approach based on collective classification (CC), which accounts for both the intrinsic and extrinsic factors by utilizing the local features of, and dependencies among, individuals during prediction steps. We evaluate our CC approach using real data provided by an established mobile social networking site, with a primary focus on prediction of churn in chat activities. Our results demonstrate that using CC and social features derived from interaction records and network structure yields substantially improved prediction in comparison to using conventional classification and user profile features only.

Knowing a Good Show When You See One

James Lanagan

Social media has become an integral part of the web, and its popularity continues to provide an outlet for people's opinions and discussion about any topic of interest. In this paper we examine the interest around a number of television series that are broadcast on a weekly basis. Our aim is to show that by observing solely the initial interactions of fans or users of a web forum, we can extrapolate the longer-term interest in particular episodes. We do so by observing the temporal dynamics of the conversation, and performing a clustering so as to judge how much time is required before reasonable conclusions can be drawn about the level of interest surrounding an episode. We find that early interaction trends have strong similarities with the overall conversation patterns.

Entrepreneurs' Networks: Size, Diversity and Composition Shaped by Cultures of Rationality and Trust

Thomas Schott, Maryam Cheraghi

The network around an entrepreneur is conceptualized as having structural properties of size, diversity and composition as network components of varying prominence in the entrepreneur's network. These properties are important by impacting entrepreneurs' performance in terms of innovation, exporting and expectations for growth. In our research-consortium Global Entrepreneurship Monitor, a sample of 35430 entrepreneurs in 42 countries reported on relations with up to 20 advisors. Cluster analysis of their relations discerned five components: a private network of advice relations with spouse, parents, other family and friends, a work-place network of boss, coworkers, starters and mentors, a professional network of accountants, lawyers, banks, investors, counselors and researchers, a market network of competitors, collaborators, suppliers and customers, and an international network of advice relations with persons abroad and persons who have come from abroad. Entrepreneurs' networking is embedded in their cultures with the dimension of trust and the dimension of traditionality versus rationality. Rationality and trust are hypothesized to enhance size and diversity of networks and prominence of the work-place network, of the professional network, of the market network and of the international network, but to reduce prominence of the private network, that is, the private network is hypothesized to be more prominent in a culture of traditionality and in a culture of distrust in others beyond family and friends. Cultural effects on networking are tested as macro-to-micro effects in two-level mixed linear models with fixed effects of national levels of rationality and trust and individual-level variables as controls and random effects of country, where the dependent variables are the properties of the networks. We find that rationality promotes diversity of networks and prominence of work-place network, professional network, market network and apparently also international network, but reduces prominence of the private network. Trust increases size of the networks, diversity of networks, prominence of market network and apparently also work-place network and professional network, but reduces prominence of the private network. These cultural effects are larger than most effects of entrepreneurs' individual characteristics.

A Game Theoretic Framework for Community Detection

Patrick J. McSweeney, Kishan Mehrotra, Jae C. Oh

The mainstream approach for community detection focuses on the optimization of a metric that measures the quality of a partition over a given network. Optimizing a global metric is akin to community assignment by a centralized decision maker. In liu of global optimization, we treat each node as a player in a hedonic game and focus on their ability to form fair and stable community structures. Application on real-world networks and a well-known benchmark demonstrates that our approach produces better results than modularity optimization.

Communities and Balance in Signed Networks: A Spectral Approach

Pranay Anchuri, Malik Magdon-Ismail

Discussion based websites like Epinions.com and Slashdot.com allow users to identify both friends and foes. Such networks are called Signed Social Networks and mining communities of like-minded users from these networks has potential value. We extend existing community detection algorithms that work only on unsigned networks to be applicable to signed networks. In particular, we develop a spectral approach augmented with iterative optimization. We use our algorithms to study both communities and structural balance. Our results indicate that modularity based communities are distinct from structurally balanced communities.

A Hybrid Evolutionary Algorithm Based on HSA and CLS for Multi-objective Community Detection in Complex Networks

Babak Amiri, Liaquat Hossain, John Crawford

Detecting community structure is crucial for uncovering the links between structures and functions in complex networks. Most of contemporary community detection algorithms employ single optimization criteria (e.g., modularity), which may have fundamental disadvantages. This paper considers the community detection process as a Multi-Objective optimization Problem (MOP). To solve the community detection problem this study used modified harmony search algorithm (HAS), the original HAS often converges to local optima which is a disadvantage with this method. To avoid this shortcoming the HAS was combined with a Chaotic Local Search (CLS). In the proposed algorithm an external repository considered to save non-dominated solutions found during the search process and a fuzzy clustering technique was used to control the size of the repository. The experiments in synthetic and real networks show that the proposed multi-objective community detection algorithm is able to discover more accurate community structures.

Optimal Spatial Resolution for the Analysis of Human Mobility

Michele Coscia, Salvatore Rinzivillo, Fosca Giannotti, Dino Pedreschi

The availability of massive network and mobility data from diverse domains has fostered the analysis of human behaviors and interactions. This data availability leads to challenges in the knowledge discovery community. Several different analyses have been performed on the traces of human trajectories, such as understanding the real borders of human mobility or mining social interactions derived from mobility and vice versa. However, the data quality of the digital traces of human mobility has a dramatic impact over the knowledge that it is possible to mine, and this issue has not been thoroughly tackled so far in literature. In this paper, we mine and analyze with complex network techniques a large dataset of human trajectories, a GPS dataset from more than 150k vehicles in Italy. We build a multi resolution grid and we map the trajectories with several complex networks, by connecting the different areas of our region of interest. Then we analyze the structural properties of these networks and the quality of the borders it is possible to infer from them. The result is a significant advancement in our understanding of the data transformation process that is needed to connect mobility with social network analysis and mining.

A New Algorithm for Positive Influence Dominating Set in Social Networks

Hassan Raei, Nasser Yazdani, Masoud Asadpour

Positive Influence Dominating Set (PIDS) has applications in Online Social Networks (OSN) such as Viral Marketing and College Drinking Problem. To many reasons finding Minimum PIDS (MPIDS) is very desirable. Beside, one of the most important features that distinguish the graph of OSN from other networks is Power-Law degree distribution. Unfortunately computing MPIDS in Power-Law graph is a NP-Complete problem. Recently, one greedy algorithm has been proposed in the literature for the PIDS problem with time complexity of $O(n^3)$. In this paper, we propose a new greedy algorithm for PIDS which has outstanding time complexity of $O(n^2)$. Theoretical analysis and simulation results are also presented to verify our approach's efficiency. The simulation results reveal that compared to other algorithm, our algorithm efficiently reduces the PIDS size.

Relative Validity Criteria for Community Mining Algorithms

Reihaneh Rabbany, Mansoreh Takaffoli, Justin Fagnan, Osmar R. Zùane, Ricardo J.G.B. Campello

Grouping data points is one of the fundamental tasks in data mining, which is commonly known as clustering if data points are described by attributes. When dealing with interrelated data that does not have any attributes and is represented in the form of nodes and their relationships, this task is also referred to as community mining. There has been a considerable number of approaches proposed in recent years for mining communities in a given network. But little work has been done on how to evaluate community mining results. The common practice is to use an agreement measure to compare the mining result against a ground truth, however, the ground truth is not known in most of the real world applications. In this paper, we investigate relative clustering quality measures defined for evaluation of clustering data points with attributes and propose proper adaptations to make them applicable in the context of social networks. Not only these relative criteria could be used as metrics for evaluating quality of the groupings but also they could be used as objectives for designing new community mining algorithms.

A Community Based Algorithm for Deriving Users' Profiles from Egocentrics Networks

Dieudonné Tchuente, Marie-Francoise Canut, Nadine Baptiste-Jessel, André Péninou, Florence Sedes

nowadays, social networks are more and more widely used as a solution for enriching users'™ profiles in systems such as recommender systems or personalized systems. For an unknown user'™s interest, the user'™s social network can be a meaningful data source for deriving that interest. However, in the literature very few techniques are designed to meet this solution. Existing techniques usually focus on people individually selected in the user'™s social network, and strongly depend on each author'™s objective. To improve these techniques, we propose to use a community based algorithm that is applied to a part of the user'™s social network (egocentric network) and that can be reused for any purpose (e.g. personalization, recommendation). We compute weighted user'™s interests from these communities by considering their semantics (interests related to communities) and their structural measures (e.g. centrality measures) in the egocentric network graph. A first experiment conducted in Facebook demonstrates the usefulness of this technique compared to individuals based techniques, and the influence of structural measures (related to communities) on the quality of derived profiles. The results also raise the problem of users'™ privacy in platforms such as online social networks. To enable users to better protect their privacy, these platforms should provide their users with a way to also make their friend list private.

Percolation Computation in Complex Networks

Fergal Reid, Aaron McDaid, Neil Hurley

K-clique percolation is an overlapping community finding algorithm which extracts particular structures, comprised of overlapping cliques, from complex networks. While it is conceptually straightforward, and can be elegantly expressed using clique graphs, certain aspects of k-clique percolation are computationally challenging in practice. In this paper we investigate aspects of empirical social networks, such as the large numbers of overlapping maximal cliques contained within them, that make clique percolation, and clique graph representations, computationally expensive. We motivate a simple algorithm to conduct clique percolation, and investigate its performance compared to current best-in-class algorithms. We present improvements to this algorithm, which allow us to perform k-clique percolation on much larger empirical datasets. Our approaches perform much better than existing algorithms on networks exhibiting pervasively overlapping community structure, especially for higher values of k. However, clique percolation remains a hard computational problem, current algorithms still scale worse than some other overlapping community finding algorithms.

Churn Prediction in a Real Online Social Network Using Local CommunIty Analysis

Blaise Ngonmang, Emmanuel Viennet, Maurice Tchuente

Prediction of user behavior in Social Networks is important for a lot of applications, ranging from marketing to social community management. In this paper, we develop and test a model to estimate the propensity of a user to stop using the social platform in a near future. This problem is called churn prediction and has been extensively studied in telecommunication networks. We focus here on building a statistical model estimating the probability that a user will leave the social network in the near future. The model is based on graph attributes extracted in the user's vicinity. We present a novel algorithm to accurately detect overlapping local communities in social graphs. Our algorithm outperforms the state of the art methods and is able to deal with pathological cases which can occur in real networks. We show that using attributes computed from the local community around the user allows to build a robust statistical model to predict churn. Our ideas are tested on one of the largest French social blog platform, Sky rock, where millions of teenagers interact daily.

Using Field of Research Codes to Discover Research Groups from Co-authorship Networks

Qinxue Meng, Paul J. Kennedy

Nowadays, academic collaboration has become more prevalent and crucial than ever before and many studies of academic collaboration analysis are implemented based on coauthor ship networks. This paper aims to build a novel coauthor ship network by importing field of research codes based on Newman's model, and then analyze and extract research groups via spectral clustering. In order to explain the effectiveness of this revised network, we take the academic collaboration at the University of Technology, Sydney (UTS) as an example. The result of this study advances methods for selecting the most prolific research groups and individuals in research institutions, and provides scientific evidence for policymakers to manage laboratories and research groups more efficiently in the future.

A Tunable Graph Model for Incorporating Geographic Spread in Social Graph Models

Rajesh Sharma, Anwitaman Datta

Modeling and understanding social network structure has interested researchers from many backgrounds including social science, computer science, theoretical physics and graph theory. Notable models include [1] and [2] achieving graphs with power-law degree distribution using preferential attachment and small-world characteristics using randomized rewiring of a regular ring lattice respectively. In contrast to a body of follow-up research which refine upon these seminal works to better capture the graph structure and characteristics (such as improving clustering coefficient by considering social triads along with preferential attachment [3]), this work aims additionally to model the geographic spread in social networks. With increased mobility in our society as well as enhanced communication opportunities social networks are increasingly spread all over the globe. Synthetic graphs imitating real-world social network characteristics are often used for driving simulations for planning and decision support. Incorporating geographic spread can facilitate better infrastructure provisioning in distributed systems supporting social and collaborative applications or model information of malware diffusion, word-of-mouth marketing, etc. The proposed model is tunable and modular. The model can be tuned to produce graphs with different geographic spread. The model is modular in the sense that existing geographic spread agnostic social network models can be plugged into our model to achieve desirable geographic spread in addition to other characteristics (such as degree distribution, clustering coefficient) that such a model would natively support.

Predicting Personality with Social Behavior

Sibel Adali, Jennifer Golbeck

In this paper, we examine to which degree behavioral measures can be used to predict personality. Personality is one factor that dictates people's propensity to trust and their relationships with others. In previous work, we have shown that personality can be predicted relatively accurately by analyzing social media profiles. We demonstrated this using public data from facebook profiles and text from Twitter streams. As social situations are crucial in the formation of one's personality, one's social behavior could be a strong indicator of her personality. Given most users of social media sites typically have a large number of friends and followers, considering only these aspects may not provide an accurate picture of personality. To overcome this problem, we develop a set of measures based on one's behavior towards her friends and followers. We introduce a number of measures that are based on the intensity and number of social interactions one has with friends along a number of dimensions such as reciprocity and priority. We analyze these features along with a set of features based on the textual analysis of the messages sent by the users. We show that behavioral features are very useful in determining personality and perform as well as textual features.

Are All Social Networks Structurally Similar?

Aneeq Hashmi, Faraz Zaidi, Arnaud Sallaberry, Tariq Mehmood

The modern age has seen an exponential growth of social network data available on the web. Analysis of these networks reveal important structural information about these networks in particular and about our societies in general. More often than not, analysis of these networks is concerned in identifying similarities among social networks and how they are different from other networks such as protein interaction networks, computer networks and food web. In this paper, our objective is to perform a critical analysis of different social networks using structural metrics in an effort to highlight their similarities and differences. We use five different social network datasets which are contextually and semantically different from each other. We then analyze these networks using a number of different network statistics and metrics. Our results show that although these social networks have been constructed from different contexts, they are structurally similar.

AuthorRank+FOAF: Ranking for Co-Authorship Networks on the Web

Lule Ahmedi

The FOAF ontology has become a core model for describing social data of people and their links on the Web, and the people ranking is growing in popularity since search engines are considering the author's reputation of a Web page when generating search results. FOAF alone is yet insufficient to model social networks for ranking people on the Web. We concentrate here in co-authorship networks as a special case of social networks, and propose a model which extends FOAF with Page Rank and Author Rank metrics, both implemented in Semantic Web Rule Language (SWRL), for gauging the reputation of authors. Preliminary results are demonstrated, showcasing also the huge potential of this ranking approach for adopting it on social networks in general, and by search engines in particular where our future work will focus.

Extracting Celebrities from Online Discussions

Mathilde Forestier, Julien Velcin, Anna Stavrianou, Djamel Zighed

Online discussions became increasingly widespread with the Web 2.0: no matter the distance, whether you know the person or not, you can discuss and exchange ideas with people all over the world through forums, blogs, and newsgroups. The news websites have extensively used forums in order to encourage the reader being a real participant in the information media. This paper aims at automatically extracting the celebrities from such discussions. We propose certain meta-criteria and we provide an evaluation on a dataset of 35,175 posts written by 14,443 users. The results show that one of the proposed meta-criteria succeeds in extracting celebrities and allows for further improvements.

Relevance of SIR Model for Real-world Spreading Phenomena: Experiments on a Large-scale P2P System

Daniel F. Bernardes, Matthieu Latapy, Fabien Tarissan

Understanding the spread of information on complex networks is a key issue from a theoretical and applied perspective. Despite the effort in developing theoretical models for this phenomenon, gauging them with large-scale real-world data remains an important challenge due to the scarcity of open, extensive and detailed data. In this paper, we explain how traces of peer-to-peer file sharing may be used to this goal. We also perform simulations to assess the relevance of the standard SIR model to mimic key properties of real spreading cascades. We examine the impact of the network topology on observed properties and finally turn to the evaluation of two heterogeneous extensions of the SIR model. We conclude that all the models tested failed to reproduce key properties of such cascades: real spreading cascades are relatively "elongated" compared to simulated ones. We have also observed some interesting similarities common to all SIR models tested.

How do Facebookers Use Friendlists

Yousra Javed, Mohamed Shehab

Facebook friend lists are used to classify friends into groups and assist users in controlling access to their information. In this paper, we study the effectiveness of Facebook friend lists from two aspects: Friend Management and Policy Patterns by examining how users build friend lists and to what extent they use them in their policy templates. We have collected real Facebook profile information and photo privacy policies of 222 participants, through their consent in our Facebook survey application posted on Mechanical Turk. Our data analysis shows that users' customized friend lists are less frequently created and have fewer overlaps as compared to Facebook created friend lists. Also, users do not place all of their friends into lists. Moreover, friends in more than one friend lists have higher values of node betweenness and outgoing to incoming edge ratio values among all the friends of a particular user. Last but not the least, friend list and user based exceptions are less frequently used in policies as compared to allowing all friends, friends of friends and everyone to view photos.

Mining User's Real Social Circle in Microblog

Hailong Qin, Ting Liu, Yanjun Ma

As a media and communication platform, microblog is more and more popular around the world. Users can follow anyone ranges from well-known individuals to real friends, and read their tweets without their permission. Most users follow a large number of celebrities and public media in microblog, however, these celebrities do not necessarily follow all their fans. Such one-way relationship abounds in the user network and is displayed in the forms of users' followees and followers, which make it difficult to identify users' real friends who are contained in the merged list of followees and followers. The aim of this paper is to propose a general algorithm for mining users' real friends in social media and dividing them into different social circles automatically according to the closeness of their relationships. To verify the effectiveness of the proposed algorithm, we build a microblog application which presents the social circles for users identified by the algorithm and enable users to modify the proposed results according to her/his real social circles. We demonstrate that our algorithm is superior to traditional clustering method in terms of F measure and Mean Average Precision.

On a Triadic Approach to Connect Microstructural Properties to Social Macrostructural Patterns

Yuxi Hu, Mina Doroud, S. Felix Wu

Social macrostructures, such as structural balance, ranked clusters and transitivity, are of great importance on account of their abilities to reflect the underlying social psychological processes about the formation and evolution of relationships among people. Here we present a detailed study on examining the existence and evolution of social macrostructures in an empirical online social network, and exploring how they can be explained by network micro structural properties, i.e. nodal in degree and out degree and dyadic feature. We establish the micro-macro linkage by analyzing the network triadic patterns. Based on a novel clustering coefficient based network sampling approach, we show that the distribution of observed triad census in our data is low dimensional and can be greatly explained by network dyadic properties. In a time series analysis, we observe that our network exhibits strong tendencies towards balanced, transitive and clustered social macrostructure given the nodal and dyadic characteristics. Our findings supplement the studies on structural properties of online social network by providing more insights on the relation between network macrostructures and the micro-level social processes that result in them. And they form the basis to understand better how online social media systems change the information and communication fabric of our society.

Semi-Supervised Policy Recommendation for Online Social Networks

Mohamed Shehab, Hakim Touati

Fine grain policy settings in social network sites is becoming a very important requirement for managing user's privacy. Incorrect privacy policy settings can easily lead to leaks in private and personal information. At the same time, being too restrictive would reduce the benefits of online social networks. This is further complicated with the growing adoption of social networks and with the rapid growth in information uploading and sharing. The problem of facilitating policy settings has attracted numerous access control, and human computer interaction researchers. The solutions proposed range from usable interfaces for policy settings to automated policy settings. We propose a fine grained policy recommendation system that is based on an iterative semi-supervised learning approach that uses the social graph propagation properties. Active learning and social graph properties were used to detect the most informative instances to be labeled as training sets. We implemented and tested our approach using real Facebook dataset. We compared our proposed approach to supervised learning and random walk approaches. Our proposed approaches provided high accuracy and precision when compared to the other approaches.

Learning the Strength of the Factors Influencing User Behavior in Online Social Networks

Bo Hu, Mohsen Jamali, Martin Ester

As social networking is moving into the web, the study and exploitation of social correlation has emerged as a hot research topic. Most of these work consider binary social relations, called "friendships". However, online users tend to establish many friendships of varying degree of strength, e.g., relatives, friends, co-workers, and acquaintances. We argue that, due to their different degree of strength, different friend relationships will have greatly varying degrees of correlation and should be distinguished. Besides, social correlation is not the only factor driving user behavior. In this paper, we address the problem of learning the strength of the social correlation, user, item, and sparsity factors in online social networks. We propose a probabilistic model, Factor Weight Model, for learning these strengths which maximize the joint probability of the observed user behavior, i.e., actions on items. Different from existing methods, our model considers not only social correlation, but it also considers the other factors affecting user behavior. We have conducted experiments on four real life data sets from Epinions, Flixster, Flickr, and Digg. Our experiments prove the superiority of our model over a state-of-the-art method in terms of action prediction. We also analyze the contributions of the various factors for the prediction performance.

Link Prediction: Fair and Effective Evaluation

Ryan Lichtnwalter, Nitesh V. Chawla

Link prediction is a popular area for publication. Papers appear in virtually every conference on data mining or network science with new methods. We argue that the practical performance potential of these methods is generally unknown because of challenges endemic to evaluation in many link prediction contexts. We demonstrate that current methods of evaluation are inadequate and can lead to woefully errant conclusions about practical performance potential. We argue for the use of precision-recall threshold curves and associated areas in lieu of receiver operating characteristic curves due to the extreme imbalance of the link prediction classification problem. We provide empirical examples of how current methods lead to questionable conclusions, how the fallacy of these conclusions is illuminated by methods we propose, and suggest a fair and consistent framework for link prediction evaluation for longitudinal and non-longitudinal network data sets.

Predicting Friends and Foes in Signed Networks Using Inductive Inference and Social Balance Theory

Arti Patidar, Vinti Agarwal, K.K. Bharadwaj

Besides the notion of friendship, trust or support in social networking sites (SNSs), quite often social interactions also reflect users' antagonistic attitude towards each other. Thus, the hidden knowledge contained in social network data can be considered as an important resource to discover the formation of such positive and negative links. In this work, an inductive learning framework is presented to suggest 'friends' and 'foes' links to individuals which envisage the social balance among users in the corresponding friends and foes networks (FFN). First we learn a model by applying C4.5, the most widely adopted decision tree based classification algorithm, to exploit the feature patterns presented in the users' FFN and utilizing it to further predict friend/foe relationship of unknown links. Secondly, a quantitative measure of social balance, balance index, is used to support our decision on the recommendation of new friends and foes links (FFL) to avoid possible imbalance in the extended FFN with newly suggested links. The proposed scheme ensures that the recommendation of new FFLs either maintains or enhances the balancing factor of the existing FFN of an individual. Experimental results show the effectiveness of our proposed schemes.

The Impact of Measurement Time on Subgroup Detection in Online Communities

Sam Zeini, Tilman Gⵔhnert, Ulrich Hoppe, Lothar Krempel

More and more communities use internet based services and infrastructure for communication and collaboration. All these activities leave digital traces that are of interest for research as real world data sources that can be processed automatically or semi-automatically. Since productive online communities (such as open source developer teams) tend to support the establishment of ties between actors who work on or communicate about the same or similar objects, social network analysis is a frequently used research methodology in this field. A typical application of SNA techniques is the detection of cohesive subgroups of actors (also called "community detection"). A relatively new method for detecting cohesive subgroups is the Clique Percolation Method (CPM), which allows for detecting overlapping subgroups. We have used CPM to analyze data from some open source developer communities (mailing lists and log files) and have compared the results for varied time windows of measurement. The influence of the time span of data capturing/aggregation can be compared to photography: A certain minimal window size is needed to get a clear image with enough "light" (i.e. dense enough interaction data), whereas for very long time spans the image will be blurred because subgroup membership will indeed change during the time span (corresponding to a moving target). In this sense, our target parameter is "resolution" of subgroup structures. We have identified several indicators for good resolution. Applying these indicators to the different CPM results shows the best resolution is a time span of around 2-3 months. In general, this value will vary for different types of communities with different communication frequency and behavior. Following our findings, an explicit

analysis and comparison of the influence of time window for different communities may be used to better adjust analysis techniques for the communities at hand.

On Learning Cluster Coefficient of Private Networks

Yue Wang, Xintao Wu, Jun Zhu, Yang Xiang

Enabling accurate analysis of social network data while preserving differential privacy has been challenging since graph features such as clustering coefficient or modularity often have high sensitivity, which is different from traditional aggregate functions (e.g., count and sum) on tabular data. In this paper, we treat a graph statistics as a function f and develop a divide and conquer approach to enforce differential privacy. The basic procedure of this approach is to first decompose the target computation f into several less complex unit computations $f1, \cdots, f_m$ connected by basic mathematical operations (e.g., addition, subtraction, multiplication, division), then perturb the output of each f_i with Lap lace noise derived from its own sensitivity value and the distributed privacy threshold ϵ_i, and finally combine those perturbed f_i as the perturbed output of computation f. We examine how various operations affect the accuracy of complex computations. When unit computations have large global sensitivity values, we enforce the differential privacy by calibrating noise based on the smooth sensitivity, rather than the global sensitivity. By doing this, we achieve the strict differential privacy guarantee with smaller magnitude noise. We illustrate our approach by using clustering coefficient, which is a popular statistics used in social network analysis. Empirical evaluations show the developed divide and conquer approach outperforms the direct approach.

Network Anomaly Detection Using Co-clustering

Evangelos E. Papalexakis, Alex Beutel, Peter Steenkiste

Early Internet architecture design goals did not put security as a high priority. However, today Internet security is a quickly growing concern. The prevalence of Internet attacks has increased significantly, but still the challenge of detecting such attacks generally falls on the end hosts and service providers, requiring system administrators to detect and block attacks on their own. In particular, as social networks have become central hubs of information and communication, they are increasingly the target of attention and attacks. This creates a challenge of carefully distinguishing malicious connections from normal ones. Previous work has shown that for a variety of Internet attacks, there is a small subset of connection measurements that are good indicators of whether a connection is part of an attack or not. In this paper we look at the effectiveness of using two different co-clustering algorithms to both cluster connections as well as mark which connection measurements are strong indicators of what makes any given cluster anomalous relative to the total data set. We run experiments with these co-clustering algorithms on the KDD 1999 Cup data set. In our experiments we find that soft co-clustering, running on samples of data, finds consistent parameters that are

strong indicators of anomalous detections and creates clusters, that are highly pure. When running hard co-clustering on the full data set (over 100 runs), we on average have one cluster with 92.44% attack connections and the other with 75.84% normal connections. These results are on par with the KDD 1999 Cup winning entry, showing that co-clustering is a strong, unsupervised method for separating normal connections from anomalous ones. Finally, we believe that the ideas presented in this work may inspire research for anomaly detection in social networks, such as identifying spammers and fraudsters.

A Guide to Differential Privacy Theory in Social Network Analysis

Christine Task, Chris Clifton

Privacy of social network data is a growing concern which threatens to limit access to this valuable data source. Analysis of the graph structure of social networks can provide valuable information for revenue generation and social science research, but unfortunately, ensuring this analysis does not violate individual privacy is difficult. Simply anonymizing graphs or even releasing only aggregate results of analysis may not provide sufficient protection. Differential privacy is an alternative privacy model, popular in data-mining over tabular data, which uses noise to obscure individuals' contributions to aggregate results and offers a very strong mathematical guarantee that individuals' presence in the data-set is hidden. Analyses that were previously vulnerable to identification of individuals and extraction of private data may be safely released under differential-privacy guarantees. We review two existing standards for adapting differential privacy to network data and analyse the feasibility of several common social-network analysis techniques under these standards. Additionally, we propose out-link privacy, a novel standard for differential privacy over network data, and introduce two powerful out-link private algorithms for common network analysis techniques that were infeasible to privatize under previous differential privacy standards.

Anonymizing Subsets of Social Networks with Degree Constrained Subgraphs

Sean Chester, Jared Gaertner, Ulrike Stege, S. Venkatesh

In recent years, concerns of privacy have become more prominent for social networks. Anonymizing a graph meaningfully is a challenging problem, as the original graph properties must be preserved as well as possible. We introduce a generalization of the degree anonymization problem posed by Liu and Terzi. In this problem, our goal is to anonymize a given subset of nodes while adding the fewest possible number of edges. The main contribution of this paper is an efficient algorithm for this problem by exploring its connection with the degree-constrained sub graph problem. Our experimental results show that our algorithm performs very well on many instances of social network data.

Privacy Preservation by k-Anonymization of Weighted Social Networks

Maria E. Skarkala, Manolis Maragoudakis, Stefanos Gritzalis, Lilian Mitrou, Hannu Toivonen, Pirjo Moen

Privacy preserving analysis of a social network aims at a better understanding of the network and its behavior, while at the same time protecting the privacy of its individuals. We propose an anonymization method for weighted graphs, i.e., for social networks where the strengths of links are important. This is in contrast with many previous studies which only consider unweighted graphs. Weights can be essential for social network analysis, but they pose new challenges to privacy preserving network analysis. In this paper, we mainly consider prevention of identity disclosure, but we also touch on edge and edge weight disclosure in weighted graphs. We propose a method that provides k-anonymity of nodes against attacks where the adversary has information about the structure of the network, including its edge weights. The method is efficient, and it has been evaluated in terms of privacy and utility on real word datasets.

PROTOSS: A Run Time Tool for Detecting Privacy Violations in Online Social Networks

Ozgür Kafali, Akin Günay, Pinar Yolum

As online social networks are becoming part of both social and work life, preserving privacy of their users is becoming tremendously difficult. While these social networks are promising privacy through privacy agreements, everyday new privacy leakages are emerging. Ideally, online social networks should be able to manage and maintain their agreements through well-founded methods. However, the dynamic nature of the online social networks is making it difficult to keep private information contained. We have developed PROTOSS, a run time tool for detecting privacy leakages in online social networks. PROTOSS captures relations among users, their privacy agreements with an online social network operator, and domain-based inference rules. It then uses model checking to detect if an online social network will leak private information.

Exploiting and Evaluating MapReduce for Large-Scale Graph Mining

Hung-Che Lai, Cheng-Te Li, Yi-Chen Lo, Shou-De Lin

Graph mining is a popular technique for discovering the hidden structures or important instances in a graph, but the computational efficiency is usually a cause for concern when dealing with large-scale graphs containing billions of entities. Cloud computing is widely regarded as a feasible solution to the problem. In this work, we present an open source graph mining library called the MapReduce Graph Mining Framework (MGMF) to be a robust and efficient MapReduce-based graph mining tool. We start from dividing graph mining algorithms into four categories and designing a MapReduce framework for algorithms in each category. The experimental results show that MGMF is 3 to 20 times more efficient

than PEGASUS, a state-of-the-art library for graph mining on MapReduce. Moreover, it provides better coverage of different graph mining algorithms. We also validate our framework on billion-scaled networks to demonstrate that it is scalable to the number of machines. Furthermore, we test and compare the feasibility between single ma-chine and the cloud computing technique. The effects of different file input formats for MapReduce are investigated as well. Our implemented open-source library can be downloaded from http://mslab.csie.ntu.edu.tw/~noahsark/MGMF/.

Link Prediction in a Modified Heterogeneous Bibliographic Network

John Boaz Lee, Henry Adorna

Researchers have discovered, in recent years, the advantages of modeling complex systems using heterogeneous information networks. These networks are comprised of heterogeneous sets of nodes and edges that better represent the different entities and relationships often found in the real world. Although heterogeneous networks provide a richer semantic view of the data, the added complexity makes it difficult to directly apply existing techniques that work well on homogeneous networks. In this paper, we propose a graph modification process that alters an existing heterogeneous bibliographic network into another network, with the purpose of highlighting the important relations in the bibliographic network. Several importance scores, some adopted from existing work and others defined in this work, are then used to measure the importance of links in the modified network. The link prediction problem is studied on the modified network by implementing a random walk-based algorithm on the network. The importance scores and the structure of the modified graph are used to guide a random walker towards relevant parts of the graph, i.e. towards nodes to which new links will be created in the future. The different properties of the proposed algorithm are evaluated experimentally on a real world bibliographic network, the DBLP. Results show that the proposed method outperforms the state-of-the-art supervised technique as well as various approaches based on topology and node attributes.

Fast Exact Computation of betweenness Centrality in Social Networks

Miriam Baglioni, Filippo Geraci, Marco Pellegrini, Ernesto Lastres

Social networks have demonstrated in the last few years to be a powerful and flexible concept useful to represent and analyze data emerging form social interactions and social activities. The study of these networks can thus provide a deeper understanding of many emergent global phenomena. The amount of data available in the form of social networks data is growing by the day, and this poses many computational challenging problems for their analysis. In fact many analysis tools suitable to analyze small to medium sized networks are inefficient for large social networks. The computation of the betweenness centrality index is a well established method for network data analysis and it is also important as subroutine in more advanced algorithms, such as the Girvan-Newman method for graph partitioning. In this paper we present a

new approach for the computation of the betweenness centrality, which speeds up considerably Brandes' algorithm (the current state of the art) in the context of social networks. Our approach exploits the natural sparsity of the data to algebraically (and efficiently) determine the betweenness of those nodes forming trees (tree-nodes) in the social network. Moreover, for the residual network, which is often of much smaller size, we modify directly the Brandes' algorithm so that we can remove the nodes already processed and perform the computation of the shortest paths only for the residual nodes. Tests conducted on a sample of publicly available large networks from the Stanford repository show that improvements of a factor ranging between 2 and 5 are possible on several such graphs, when the sparsity, measured by the ratio of tree-nodes to the total number of nodes, is in a medium range (30% to 50%). For some large networks from the Stanford repository and for a sample of social networks provided by Sistemi Territoriali with high sparsity (80% and above) tests show that our algorithm consistently runs between one and two orders of magnitude faster than the current state of the art exact algorithm.

Enhancing Academic Event Participation with Context-aware and Social Recommendations

Manh Cuong Pham, Dejan Kovachev, Yiwei Cao, Ghislain Manib Mbogos, Ralf Klamma

The plethora of talks and presentations taking place at academic conferences makes it difficult, especially for young researchers to attend the right talks or discuss with participants and potential collaborators with similar interests. Participants may not have a priori knowledge that allows them to select the right talks or informal interactions with other participants. In this paper we present the context-aware mobile recommendation services (CAMRS) based on the current context (whereabouts at the venue, popularity and activities of talks and presentations) sensed at the conference venue. Additionally, we augment the current context with the academic community context of conference participants which is inferred by using social network analysis and link prediction on large-scale co-authorship and citation networks of participants. By combining the dynamic and social context of participants, we are able to recommend talks and people that may be interesting to a particular participant. We evaluated CAMRS using data from two large digital libraries - the DBLP and CiteSeerX, and participants from two conferences - ICWL 2010 and EC-TEL 2011. The result shows that the new approach can recommend novel talks and helps participants in establishing new connections at conference venue.

Learning Rating Patterns for Top-N Recommendations

Yongli Ren, Gang Li, Wanlei Zhou

Two rating patterns exist in the user-item rating matrix and influence each other: the personal rating patterns are hidden in each user's entire rating history, while the global rating patterns are hidden in the entire user-item rating matrix. In this paper, a Rating Pattern Subspace is proposed to model both of the rating patterns simultaneously by iteratively refining each other with an EM-like algorithm. Firstly, a low-rank

subspace is built up to model the global rating patterns from the whole user-item rating matrix, then, the projection for each user on the subspace is refined individually based on his/her own entire rating history. After that, the refined user projections on the subspace are used to improve the modelling of the global rating patterns. Iteratively, we can obtain a well-trained low-rank Rating Pattern Subspace, which is capable of modelling both the personal and the global rating patterns. Based on this subspace, we propose a RapSVD algorithm to generate Top-N recommendations, and the experiment results show that the proposed method can significantly outperform the other state-of-the-art Top-N recommendation methods in terms of accuracy, especially on long tail item recommendations.

Fast Recommendation on Bibliographic Networks

Onur Küçüktunç, Kamer Kaya, Erik Saule, mit V. ⎮atalyürek

Graphs and matrices are widely used in algorithms for social network analyses. Since the number of interactions is much less than the possible number of interactions, the graphs and matrices used in the analyses are usually sparse. In this paper, we propose an efficient implementation of a sparse-matrix computation which arises in our publicly available citation recommendation service called the advisor. The recommendation algorithm uses a sparse matrix generated from the citation graph. We observed that the nonzero pattern of this matrix is highly irregular and the computation suffers from high number of cache misses. We propose techniques for storing the matrix in memory efficiently and reducing the number of cache misses. Experimental results show that our techniques are highly efficient on reducing the query processing time which is highly crucial for a web service.

Personalization with Dynamic Group Profile

Kamal Taha, Ramez Elmasri

In this paper, we propose an XML-based recommender system, called PDGP. It is a type of collaborative information filtering system. PDGP uses ontology-driven social networks, where nodes represent social groups. A social group is an entity that defines a group based on demographic, ethnic, cultural, religious, age, or other characteristics. In the PDGP framework, query results are filtered and ranked based on the preferences of the social groups to which the user belongs. The user's social groups are inferred implicitly by the system without involving the user. PDGP constructs the social groups and identifies their preferences dynamically on the fly. These preferences are determined from the preferences of the social groups' member users using a group modeling strategy. PDGP can be used for various practical applications, such as Internet or other businesses that market preference-driven products. We experimentally compared PDGP with an existing system. Results showed marked improvement.

Dealing with Disappearance of an Actor Set in Social Networks

Idrissa Sarr, Rokia Missaoui, Romain Lalande

Social networks are dynamic structures that contain a set of entities and links. In such a dynamic environment, a specific node or a group of nodes can play an important role in the information flow transmission within the network and therefore, its disappearance may lead to a disconnected network or a breakdown in the information flow. The objective of this paper is to extend our previous work on managing a node disappearance to handling the disappearance of a group of nodes. The proposed approach relies on the role played by a group of nodes to conduct network changes and maintain the network connected while restoring the information flow with a similar quality as before the group disappearance. We consider two situations (only one versus many communities) and categorize groups of nodes into three classes (scattered, contiguous and hybrid). Hence, we manage a group disappearance with respect to its class and the network topology by adding new links in a parsimonious way and finding a substitute for a leaving group. Our approach differs from existing link prediction solutions by the fact that it uses the information flow quality as a key performance indicator to identify the potential links to add and/or the possible substitute to a disappearing group. We implement a prototype by using an open source social network analysis library (NetworkX) and we validate our solution through experiments. The results show the benefits of our solution in terms of response time and the number of added links.

OCTracker: A Density-Based Framework for Tracking the Evolution of Overlapping Communities in OSNs

Sajid Yousuf Bhat, Muhammad Abulaish

In this paper, we propose a unified framework OCTracker for tracking overlapping community evolution in online social networks. OCTracker adapts a preliminary community structure towards dynamic changes in social networks using a novel density-based approach for detecting overlapping community structures and automatically detects evolutionary events like birth, growth, contraction, merge, split, and death of communities with time. Unlike other density-based community detection methods, the proposed method does not require the neighborhood threshold parameter to be set by the users, rather it automatically determines the same for each node locally.

Crawling Social Internetworking Systems

Francesco Buccafurri, Gianluca Lax, Antonino Nocera, Domenico Ursino

In new generation social networks, we expect that the paradigm of Social Internetworking Systems (SISs, for short) will be more and more important. In this new scenario, the role of Social Network Analysis is of course still crucial but the preliminary step to do is designing a good way

to crawl the underlying graph. While this aspect has been deeply investigated in the field of social networks, it is an open issue when moving towards SISs. Indeed, we cannot expect that a crawling strategy which is good for social networks, is still valid in a Social Internetworking Scenario, due to its specific topological features. In this paper, we first confirm the above claim and, then, define a new crawling strategy specifically conceived for SISs. Finally, we show that it fully overcomes the drawbacks of the state-of-the-art crawling strategies.

Dissemination Patterns and Associated Network Effects of Sentiments in Social Networks

Robert Hillmann, Matthias Trier

Communication in online social networks has been analyzed for some time regarding the expression of sentiments. So far, very little is known about the relationship between sentiments and network emergence, dissemination patterns and possible differences between positive and negative sentiments. The dissemination patterns analyzed in this study consist of network motifs based on triples of actors and the ties among them. These motifs are associated with common social network effects to derive meaningful insights about dissemination activities. The data basis includes several thousand social networks with textual messages classified according to embedded positive and negative sentiments. Based on this data, sub-networks are extracted and analyzed with a dynamic network motif analysis to determine dissemination patterns and associated network effects. Results indicate that the emergence of digital social networks exhibits a strong tendency towards reciprocity, followed by the dominance of hierarchy as an intermediate step leading to social clustering with hubs and transitivity effects for both positive and negative sentiments to the same extend. Sentiments embedded in exchanged textual messages do only play a secondary role in network emergence and do not express differences regarding the emergence of network patterns.

Prediction of Arrival of Nodes in a Scale Free Network

S.M. Vijay Mahantesh, Sudarshan Iyengar, M. Vijesh, Shruthi R. Nayak, NIkitha Shenoy

Most of the networks observed in real life obey power-law degree distribution. It is hypothesized that the emergence of such a degree distribution is due to preferential attachment of the nodes. Barabasi-Albert model is a generative procedure that uses preferential attachment based on degree and one can use this model to generate networks with power-law degree distribution. In this model, the network is assumed to grow one node every time step. After the evolution of such a network, it is impossible for one to predict the exact order of node arrivals. We present in this article, a novel strategy to partially predict the order of node arrivals in such an evolved network. We show that our proposed method outperforms other centrality measure based approaches. We bin the nodes and predict the order of node arrivals between the bins with an accuracy of above 80%.

Static and Dynamic Aspects of Scientific Collaboration Networks

Christian Staudt, Andrea Schumm, Henning Meyerhenke, Robert Gȯrke, Dorothea Wagner

Collaboration networks arise when we map the connections between scientists which are formed through joint publications. These networks thus display the social structure of academia, and also allow conclusions about the structure of scientific knowledge. Using the computer science publication database DBLP, we compile relations between authors and publications as graphs and proceed with examining and quantifying collaborative relations with graph-based methods. We review standard properties of the network and rank authors and publications by centrality. Additionally, we detect communities with modularity-based clustering and compare the resulting clusters to a ground-truth based on conferences and thus topical similarity. In a second part, we are the first to combine DBLP network data with data from the Dagstuhl Seminars: We investigate whether seminars of this kind, as social and academic events designed to connect researchers, leave a visible track in the structure of the collaboration network. Our results suggest that such single events are not influential enough to change the network structure significantly. However, the network structure seems to influence a participant's decision to accept or decline an invitation.

Outskewer: Using Skewness to Spot Outliers in Samples and Time Series

Sébastien Heymann, Matthieu Latapy, Clémence Magnien

Finding outliers in datasets is a classical problem of high interest for (dynamic) social network analysis. However, most methods rely on assumptions which are rarely met in practice, such as prior knowledge of some outliers or about normal behavior. We propose here Out skewer, a new approach based on the notion of skewness (a measure of the symmetry of a distribution) and its evolution when extremal values are removed one by one. Our method is easy to set up, it requires no prior knowledge on the system, and it may be used on-line. We illustrate its performance on two data sets representative of many use-cases: evolution of ego-centered views of the internet topology, and logs of queries entered into a search engine.

An Analysis of Query Forwarding Strategies for Secure and Privacy-Preserving Social Networks

Michael Dürr, Marco Maier, Kevin Wiesner

Decentralized Online Social Networks (OSNs) attempt to improve user privacy and security. One example is Vegas, a Peer-to-Peer (P2P) OSN which attempts to bring its users back into complete control of the data they share. Due to its decentralized characteristics, P2P OSNs cannot support social search functions in the same way users of centralized OSNs like Facebook are familiar with. Well-known and efficient P2P search

algorithms cannot always be applied as knowledge about the structure of the social graph can be very limited. In this paper, we present an in-depth analysis of forwarding strategies to enable social search for secure and privacy preserving P2P OSNs. We compare well-known metrics from the field of unstructured P2P networks with metrics from the area of social network analysis and evaluate their applicability for P2P OSNs like Vegas. We simulate all metrics on four distinct datasets which were generated artificially from the ER- and the BA-model and from crawling data of Lastfm and Flickr. Our evaluation shows that prioritization based on knowledge from the ego network often yields the best results.

STUN: Spatio-Temporal Uncertain (Social) Networks

Chanhyun Kang, Andrea Pugliese, John Grant, V.S. Subrahmanian

STUN is an extension of social networks in which the edges are characterized by spatio-temporal annotations, as well as uncertainty allowing us to express not only relationships between vertices, but when and where these relationships were true, and how certain we are that the relationships hold. We propose a STUN query language that consists of sub graphs with spatio-temporal constraints and uncertainty requirements. We then develop an index structure to store STUN graphs, together with an algorithm to answer such queries. We describe experiments with a real-world YouTube social network data set and show that our algorithm performs well on graphs with over a million edges.

Recurrent Structural Motifs Reflect Characteristics of Distinct Networks

Chen-Hsiang Yeang, Liang-Cheng Huang, Wei-Chung Liu

In large-scale networks, certain topological patterns may occur more frequently than expected from a null model that preserves global (such as the density of the graph) and local (such as the connectivity of each node) properties of the graph. These network motifs are the building blocks of large-scale networks and may confer functional/mechanistic implications of their underlying processes. Despite active investigations and rich literature in systems biology, network motifs are less explored in social network studies. In this work, we modified and improved the method from Milo et al. 2002 to detect significantly enriched motifs in both directed and undirected networks. We applied this method to identify 3-node and 4-node motifs from the datasets of 18 networks (4 directed and 14 undirected) covering social interactions, co-authorships, web document hyperlinks, neuronal circuitry, protein-protein interactions (PPI), trophic relations in a food web, and others. Presence and absence of enriched motifs provide rich information regarding each type of network relations. In undirected networks, triangles are enriched in almost all datasets, suggesting the prevalence of transitivity in diverse networks. However, 4-node structures lacking transitivity – diamonds and stars – are also enriched in the majority of undirected networks. In directed networks, variations of feed-forward loops are over-represented in the networks of web document and political web log hyperlinks as well as neuronal connections. In contrast, the food web is

enriched with unidirectional motifs with distinct trophic levels. These results reveal the nature of distinct types of networks and invite further explorations on the relations of network structures and types of relations.

Diffusion Centrality in Social Networks

Chanhyun Kang, Cristian Molinaro, Sarit Kraus, Yuval Shavitt, V.S. Subrahmanian

Though centrality of vertices in social networks has been extensively studied, all past efforts assume that centrality of a vertex solely depends on the structural properties of graphs. However, with the emergence of online "semantic" social networks where vertices have properties (e.g. gender, age, and other demographic data) and edges are labeled with relationships (e.g. friend, follows) and weights (measuring the strength of a relationship), it is essential that we take semantics into account when measuring centrality. Moreover, the centrality of a vertex should be tied to a diffusive property in the network - a Twitter vertex may have high centrality w.r.t. jazz, but low centrality w.r.t. Republican politics. In this paper, we propose a new notion of diffusion centrality (DC) in which semantic aspects of the graph, as well as a diffusion model of how a diffusive property p is spreading, are used to characterize the centrality of vertices. We present a hyper graph based algorithm to compute DC and report on a prototype implementation and experiments showing how we can compute DCs (using real YouTube data) on social networks in a reasonable amount of time. We compare DC with classical centrality measures like degree, closeness, betweenness, eigenvector and stress centrality and show that in all cases, DC produces higher quality results. DC is also often faster to compute than both betweenness, closeness and stress centrality, but slower than degree and eigenvector centrality.

Investigating Organized Crime Groups: A Social Network Analysis Perspective

Mohammad A. Tayebi, Uwe Glasser

In this paper, we analyze co-offending networks derived from a large real-world crime dataset for the purpose of identifying organized crime structures and their constituent entities. We focus on methodical and analytical aspects in using social network analysis methods and data mining techniques. The goal of our work is to promote computational co-offending network analysis as an effective means for extracting information about criminal organizations from large real-life crime datasets, specifically police-reported crime data. We contend that it would be virtually impossible to obtain such information by using traditional crime analysis methods. For our approach we provide an experimental evaluation with promising results.

A Semantic Triplet Based Story Classifier

Betul Ceran, Ravi Karad, Ajay Mandvekar, Steven R. Corman, Hasan Davulcu

A story is defined as "an actor(s) taking action(s) that culminates in a resolution(s)." In this paper, we investigate the utility of standard keyword based features, statistical features based on shallow-parsing (such as density of POS tags and named entities), and a new set of semantic features to develop a story classifier. This classifier is trained to identify a paragraph as a "story," if the paragraph contains mostly story(ies). Training data is a collection of expert-coded story and non-story paragraphs from RSS feeds from a list of extremist web sites. Our proposed semantic features are based on suitable aggregation and generalization of $$ triplets that can be extracted using a parser. Experimental results show that a model of statistical features alongside memory-based semantic linguistic features achieves the best accuracy with a Support Vector Machine (SVM) based classifier.

Detecting Criminal Networks Using Social Similarity

Fatih Ozgul, Zeki Erdem

Existing literature shows that social demographics features of criminal network members are important. Examples include similarity on kinship, coming from the same family, the same ethnic origin or hometown, and living in the same neighborhoods. This paper investigates whether these social similarity features can be used for detecting members of criminal networks. We developed XSDM (Extended Social Detection Model), which removes some of the weaknesses of its predecessor SODM (Social Detection Model) by adding the attribute of living in the same neighborhood in addition to having the same surname and coming from the same hometown. XSDM is tested on the Diyarbakir dataset, containing 221 drug dealing networks. XSDM detected 81 out of 221 drug dealing networks using social demographic features of individual criminals. XSDM is evaluated by recall and precision values which performed better its predecessor SODM.

The Activation of Core Social Networks in the Wake of the 22 July Oslo Bombing

Pål Roe Sundsøy, Johannes Bjelland, Geoffrey Canright, Kenth Engö-Monsen, Rich Ling

This paper examines how core social networks were activated in the wake of the bombing in Oslo on July 22, 2011. Empirical mobile traffic data illuminate exceptional behavior, just after the bombing in Oslo. We find that in the minutes after the bombing people called ties that were (1) close socially and (2) perceived to be in danger, that is, people who were close to the bombing point. Our main findings: (1) individuals first focus on their single closest contact (‘best friend'), but, soon after, switch to spending more mobile communication resources than average on

contacts ranked 2—5. (2) we see clearly a large increase (over typical) in traffic to and from, and not least within, the affected area (Oslo)—in some cases more than a 300% increase, right after the bombing. Interestingly, we also find a marked increase in traffic for relationships where both persons were outside of Oslo. All these results illustrate the importance of social contact in this highly unusual situation. This paper underscores how the mobile phone is an instrument of the intimate sphere. The situation on the 22nd of July in Oslo is a prime example of an unexpected situation, where individuals use the mobile to get critical information on their loved ones in their core network.

An Approach for the Blockmodeling in Multi-Relational Networks

Andreas Harrer, Alona Schmidt

In this paper we introduce a direct approach for Block modeling in multi-relational networks inspired by the Pajek-Approach. In this article the direct Block modeling-method is presented for two-relational networks and evaluated statistically compared to indirect approaches. In addition we apply the method to the "Krackhardt`s High-tech Managers" dataset to show the feasibility of the approach and present a different interpretation proposal for this well-known data set.

One-mode Projection of Multiplex Bipartite Graphs

Emoke-Ágnes Horvát, Katharina A. Zweig

Several important social network data sets have an inherent bipartite structure: for example, agents are affiliated with societies, authors write articles, customers buy, rent, or rate products. One commonly used network analytic approach to their analysis involves projecting them, i.e., deducing relations between actors of the same type (e.g. societies, articles, or products). Some of the available large scale data sets not only represent one, but several distinct relations between the same actors thereby calling for a projection method that accounts for the multiple nature of the relations. In this article we present a statistical method that properly extends a projection algorithm developed for bipartite networks containing one single type of relation. We show the stability of the proposed method on synthetic data. Then, we apply it to a real-world network of users rating films, namely a subset of the Netflix prize data set. We show that there is a gain from differentiating between the relation types. Based on the assumption that co-ratings of films contain information about the films' similarity, we analyze the co-liking and co-disliking structures obtained by the new one-mode projection. We find that the projections of concordant ratings show a high clustering coefficient while discordant co-ratings have a very small one. This result indicates that the assumption is valid and that thus the new one-mode projection can be used as basis for recommendations.

A Domain Specific Language Approach for Agent-Based Social Network Modeling

Enrico Franchi

Although in the past twenty years agent-based modeling has been widely adopted as a research tool in the fields of social and political sciences, there is lack of software instruments specifically created for social network simulations. Restricting the field of interest specifically to social network models and simulations instead of supporting general agent-based ones, allows for the creation of easier to use, more focused tools. In this work, we propose PyNetSym, an agent-based modeling framework designed to be friendly to programmers and non-programmers alike. PyNetSym provides a domain-specific language to specify social network simulations expressed as agent-based models. \pynetsym was created to deal with large simulations and to work effortlessly with other social network analysis toolkits.

Capability-Weighted Group Utility Maximizer for Network Coalitional Games under Uncertainty

Usha Sridhar, Sridhar Mandyam

In this paper we study network games where agents with different skills come together to cooperate and yet competitively pursue individual goals. We propose a multi-agent based utilitarian approach to model the payoff allocation problem for a class of such games where the capabilities of the agents and the payoffs are not known with certainty. The primary objective is to maximize a linear sum of the expected utilities of risk-averse agents, and we consider constant risk-aversion with exponential utility functions. We pose the problem as a stochastic cooperative game which is solved in two phases. In the first phase we apply a learning mechanism on this 'social' network of fully connected agents to arrive at a consensus on the capability of every agent in the coalition thus resolving uncertainty in capabilities. Agents initially start with a social influence matrix reflecting the influence agents have on each other and prior subjective beliefs of the capabilities of the others and these beliefs evolve through a process of interaction. We use a variant of the DeGroot algorithm to show that over time learning results in a dynamic update of the beliefs and the social influence matrix leading to a consensus. We provide theoretical convergence proofs for the algorithm. The second phase involves optimizing a capability-weighted sum of the expected utilities of the agents to achieve a group Pareto optimal solution. In this paper we propose a new framework called the Capability Weighted Group Utility Maximizer developed around Borch's theorem borrowed from the actuarial world of insurance to obtain a fair distribution of the stochastic payoffs once a consensus is reached on the capabilities of the agents in the coalition.

Sociability vs Network Dynamics: Impact of Two Aspects of Human Behavior on Diffusion Phenomena

Erick Stattner, Martine Collard, Nicolas Vidot

In this work, we focus on the issue of diffusion in dynamic networks. We begin by proposing a generic framework that merges two essential aspects of the problem: diffusion and network dynamics. Then, our framework is used for comparing effects of two aspects of the human behavior, sociability and network dynamics. On the one hand the dynamics of the network is related to the movements of individuals in a social space that lead ephemeral social proximity contacts. On the other hand the diffusion phenomenon is studied through the epidemic paradigm. The results obtained by simulations provide original insight about how and why the "social dynamics" of individuals, related to the proximity in a social space and its evolutions, influences diffusion phenomena through a population.

Detecting Probabilistic Community with Topic Modeling on Sampling SubGraphs

Zeng Feng Zeng, Bin Wu

Detecting communities plays a great important role in sociology, biology and computer science, disciplines where systems are often modeled as graphs. Such inherent community structures make us deeply understand about the networks and therefore have drawn significant interests among researchers. This paper describes a probabilistic community detection algorithm by modeling topic on sampling sub graphs. In this algorithm, the communities are modeled as latent topic variables of an LDA topic model and the vertices of sampling sub graphs are drawn from these topics with different probabilities. This paper also proposes a sub graph sampling algorithm and explores its impact on community detection performance. Our algorithm is evaluated by extensive experiments using many computer-generated artificial graphs and real-world networks. The results show that our algorithm is effective in detecting probabilistic community.

On Finding Fine-Granularity User Communities by Profile Decomposition

Seulki Lee, Minsam Ko, Keejun Han, Jae-Gil Lee

The social network represents various relationships between users, and community discovery is one of the most popular tasks analyzing these relationships. The relationships are either explicit (e.g., friends) or implicit, and we focus on community discovery with implicit relationships. Here, the key issue is how to extract the relationships between users. A user is typically represented by his/her profile, and the similarity between user profiles is measured. In most algorithms, a user has a single profile aggregating all the information about the user. For example, a profile for a researcher is a list of papers he/she wrote. This setting, however, oversimplifies the multiple characteristics of a man since individual

characteristics are mixed up. In this paper, we propose the notion and method of profile decomposition, which divides a profile into a set of sub-profiles so that they represent individual characteristics precisely. Then, we develop a community discovery algorithm, which we call DecompClus, based on profile decomposition. Using a real data set of CiteULike, we show that our proposed algorithm can precisely distinguish multiple research interests of a user and discover communities corresponding to each interest, whereas previous algorithms cannot. Overall, profile decomposition enables us to find fine-granularity user communities, thus improving the accuracy of community discovery.

Understanding Group Dynamics in Health Forums

Steven P. Crain, Shuang-Hong Yang, Hongyuan Zha

Health communities play an important role in helping patients cope with chronic disease. We analyze the vitality of these groups, in terms of their ability to attract new members and foster discussion. We present methods for performing and interpreting event history analysis with many dynamic features that are only partially observed and are highly correlated, which arise from censoring to protect privacy and the natural growth of event-count-based features respectively. We apply this methodology to identify the factors that contribute to group vitality in a diabetes community. Our findings suggest that uniformly advertising popular groups was detrimental to the diversity of popular groups, limiting the growth of the overall community. We also identified three different modes of behavior for long term members and a strategic opportunity for community managers to recruit them for leadership in less popular groups.

Identifying Long Lived Social Communities Using Structural Properties

Mark Goldberg, Malik Magdon-Ismail, James Thompson

We present a two step procedure to identify long lasting communities, or evolutions, in social networks. First, we use axiomatic foundations to `rigorously' establish shorter, strongly-connected evolutions. In the second step, we use heuristics to combine these shorter evolutions to form longer evolutions. We apply the procedure on data generated from two networks - the DBLP co-authorship database and Live Journal blog data. We visually validate our algorithms by examining the topic evolution of the associated documents. Our results demonstrate that our algorithms, based solely on structural properties of the data (who interacts with whom), are able to track thematic trends in the literature. We then use a machine learning framework to identify the structural features of the early stages of a community's evolution are most useful for predicting the lifetime of the community. We find that (in order) size, intensity and stability are the most important features.

Six Degrees of Separation among US Researchers

Hakan Kardes, Abdullah Sevincer, Mehmet Hadi Gunes, Murat Yuksel

Funding from the government agencies has been the driving force for the research and educational institutions particularly in the United States. The government funds billions of dollars every year to lead research initiatives that will shape the future. In this paper, we analyze the funds distributed by the National Science Foundation (NSF), a major source of research funding in the States, to understand the collaboration patterns among researchers and institutions. Using complex network analysis, we interpret the collaboration patterns at researcher, institution and state levels by constructing the corresponding networks based on the number of grants collaborated. We further analyze the directorates to identify the differences in collaboration trends between disciplines.

Modeling Social Network Interaction Graphs

Michael Dürr, Valentin Protschky, Claudia Linnhoff-Popien

The evaluation of novel algorithms, protocols, applications, or security attacks in context of Online Social Networks (OSN) necessitates datasets that represent a realistic snapshot of the underlying social graph. As crawling social graphs can become a time and resource consuming task, only a few anonymized datasets exist which are shared among the research community. Besides concerns about de-anonymization attacks on crawled graphs and the fact that such datasets cannot satisfy the statistical confidence in simulation results, more and more secure and privacy-preserving Peer-to-Peer (P2P) OSN architectures emerge that do not facilitate crawling of social graph data at all. In order to evaluate new metrics for OSNs in general, we need social graph models which enable the generation of synthetic datasets. In this paper we present a generic model to synthesize social interaction graphs for both centralized OSNs like Facebook and secure and privacy-preserving P2P OSNs such as Vegas. Our approach accounts for a static component which models relationships and network effects and a dynamic component which models interactions among users. A flexible parameterization schema allows our model to individually influence certain graph characteristics like node degrees, clustering coefficients, and node interactions.

Finding the Mule in the Network

John L. Pfaltz

There exist a variety of procedures for identifying clusters in large networks. This paper focuses on finding the connections between such clusters. We employ the concept of closed sets to reduce a network down to its fundamental cycles. These cycles begin to capture the global

structure of the network by eliminating a great deal of the fine detail. Nevertheless, the reduced version is completely faithful to the original. No connection in the reduced version exists unless it was in the original network, connectivity is preserved. Reduction of as much as 80% can be observed in real networks. Just reducing the size makes comprehension of the network much easier.

How Well-Connected Individuals Help Spread Influences – Analyses Based on Preferential Voter Model

Zhuo Qi Lee, Wen-Jing Hsu, Miao Lin

The spread of influence in a network is a fundamental issue in many complex systems. For instance, the word of-mouth effect in a social network is crucial to the emerging viral market. The process was previously analyzed with a voter model where an individual weighs the opinions of his acquaintances equally. However, one would expect that an individual's opinion is often more swayed by her/his better connected peers. As such, the influence of a node x on another node should be proportional to x's degree. This paper studies the spread of influence under the latter model– called Preferential Voter Model (PVM) for easier reference. We first present the exact form of the random walk stationary distribution for PVM, based on which, we show that the First Passage Time (FPT) of PVM follows an exponential decay. Furthermore, compared against the conventional (nonpreferential) voter model, we show that nodes with larger degrees exhibit faster decay rate and lower mean First Passage Time in PVM. These new results have applications to network-based complex systems. For instance, in the context of viral marketing in a social network, picking nodes of larger degrees as the starting nodes of a sales campaign not only maximizes the expected number of influenced nodes in the network, but also reduces the expected time to spread the influence. Thus, our result confirms analytically that well-connected individuals indeed exert faster and more effective influences in social networks under the preferential model.

Influence of the Dynamic Social Network Timeframe Type and Size on the Group Evolution Discovery

Stanislaw Saganowski, Piotr Brȯdka, Przemyslaw Kazienko

New technologies allow to store vast amount of data about users interaction. From those data the social network can be created. Additionally, because usually also time and dates of this activities are stored, the dynamic of such network can be analyzed by splitting it into many timeframes representing the state of the network during specific period of time. One of the most interesting issue is group evolution over time. To track group evolution the GED method can be used. However, choice of the timeframe type and length might have great influence on the method results. Therefore, in this paper, the influence of timeframe type as well as timeframe length on the GED method results is extensively analyzed.

Structural and Message Based Private Friend Recommendation

Bharath K. Samanthula, Wei Jiang

The emerging growth of online social networks have opened new doors for various business applications such as promoting a new product across its customers. Besides this, friend recommendation is an important tool for recommending potential candidates as friends to users in order to enhance the development of the entire network structure. Existing friend recommendation methods utilize social network structure and/or user profile information. However, these techniques can no longer be applicable if the privacy of users is taken into consideration. In this paper, we propose a two-phase private friend recommendation protocol for recommending friends to a given target user based on the network structure as well as utilizing the real message interaction between users. Our protocol computes the recommendation scores of all users who are within a radius of h from the target user in a privacy preserving manner. In addition, we show the practical applicability of our approach through empirical analysis.

Interactively and Visually Exploring Tours of Marked Nodes in Large Graphs

Duen Horng Chau, Leman Akoglu, Jilles Vreeken, Hanghang Tong, Christos Faloutsos

We present TOURVIZ, a interactive system for visualizing and making sense of large network datasets. Given a set of user-specified nodes of interest, TOURVIZ integrates 1) novel algorithms to find the best sub graphs that succinctly connect these nodes, and 2) visualization and interaction features that help people explore such sub graphs. We will demonstrate TOURVIZ's usage and benefits using the DBLP co-authorship graph, which consists of 329K authors (nodes) and 1094K co-authorship relations (edges). TOURVIZ can work with any kinds of graphs. We will engage the audience to try our system and comment on its usability, usefulness, and how our system may help with their work and data analytics in their domains.

A Complex Network Analysis of the United States Air Transportation

Dorothy P. Cheung, Mehmet Hadi Gunes

The air transportation industry has a great impact on the economy. In this paper, we analyze the air transportation network in the U.S. to better understand its characteristics. For this, we measure several complex network features including average shortest path, degree distribution, assortative mixing, clustering, betweenness centrality, and resiliency. We also analyzed changes over the previous two decades and compared the networks with earlier air transportation network studies. In our analysis, we did not observe considerable changes in the analyzed complex

network features of the U.S. air transportation over the past two decade. Moreover, the U.S. air transportation network exhibits small world characteristics and has an assortative mixing by the degree similar to the earlier study on the world-wide air transportation network. However, the US air transportation network exhibits a partial power law degree distribution unlike the complete power law reported in earlier air transportation networks.

Video Game Industry as a Social Network

Tony Morelli, Mehmet Hadi Gunes

The video game market is a complex system where the relationships between game titles and publishers is convoluted. While some publishers are responsible for a large number of titles others seem to be less successful. In an attempt to understand the characteristics and evolution of the video game industry, we analyze the market using social network analysis. For this purpose, we build a heterogenous network of the video game industry at different levels. Moreover, to interpret the behavior of the generations of game consoles, we build these networks for each console, i.e., PS3, Xbox 360 and Wii, and analyze the three generations they went through. We perform a comparative temporal analysis of the evolution of the industry by investigating 14 video game consoles as five generations. Finally, we try to provide insights and future directions regarding the video game market.

Building a Data Warehouse for Twitter Stream Exploration

Nafees Ur Rehman, Svetlana Mansmann, Andreas Weiler, Marc H. Scholl

In the recent year Twitter has evolved into an extremely popular social network and has revolutionized the ways of interacting and exchanging information on the Internet. By making its public stream available through a set of APIs Twitter has triggered a wave of research initiatives aimed at analysis and knowledge discovery from the data about its users and their messaging activities. While most of the projects and tools are tailored towards solving specific tasks, we pursue a goal of providing an application-independent and universal analytical platform for supporting any kind of analysis and knowledge discovery. We employ the well established data warehousing technology with its underlying multidimensional data model, ETL routine for loading and consolidating data from different sources, OLAP functionality for exploring the data and data mining tools for more sophisticated analysis. In this work we describe the process of transforming the original stream into a set of related multidimensional cubes and demonstrate how the resulting data warehouse can be used for solving a variety of analytical tasks. We expect our proposed approach to be applicable for analyzing the data of other social networks as well.

Community Detection in Social Networks Using Information Diffusion

Alireza Hajibagheri, Hamidreza Alvari, Ali Hamzeh, Sattar Hashemi

Discovering communities in popular social networks like Facebook has been receiving significant attentions recently. In this paper, inspired from real life, we have addressed the community detection problem by a framework based on Information Diffusion Model and Game Theory. In this approach, we consider each node of the social network as a selfish agent which has interactions with its neighbors and tries to maximize its total utility (i.e. received information). Finally community structure of the graph reveals after reaching to the local Nash equilibrium of the game. Experimental results on the benchmark social media datasets, synthetic and real world graphs demonstrate that our method is superior compared with the other state-of-the-art methods.

Layout Algorithm for Clustered Graphs to Analyze Community Interactions in Social Networks

Juan David Cruz, Cécile Bothorel, François Poulet

Most of the layout algorithms for clustered graphs have been designed to differentiate the groups within the graph, however they do not take into account the interactions between such groups. Identifying these interactions allows to understand how the different communities exchange messages or information, and allows the social network researcher to identify key actors, social roles and paths from one community to another. These interactions are performed by actors who actually share links with different communities. Thus, dividing the set of nodes into those who interact with several communities and those who only interact with their own communities, will give some insight about how the communities interact.

A Novel Framework for Spammer Detection in Social Bookmarking Systems

Soghra M. Gargari, Sule Gündüz Ogüdücü

Social Bookmarking systems enable users to store, organize and search their resources. Furthermore, a social book marking system allows users to share their resources with others and even join groups of people with similar interests. The data size in social book marking systems has been increased sharply in recent years with the usage of such systems. However, such systems attract spammers due to their ease of use and popularity. Spammers have started misleading search engines and other book marking system users in order to direct web traffic towards their own pages. Strong prevention and detection methods in social book marking systems are indispensable in order to stop spam activities and guaranty the accuracy and reliability of information. In this paper, we introduce a novel framework for spam detection task in social book

marking systems. Here, we propose a set of new features to improve the accuracy of spammer detection. Our experiments show that our features demonstrate a high discriminative power. A performance evaluation of our proposed method over different spammer detection methods indicate that the proposed framework yields an improvement of the prediction accuracy.

User Interests Modeling in Online Forums

Na Ni, Yaodong Li

This paper studies the problem of user modeling in online forums from a personality viewpoint. A novel hierarchical user profiling mechanism is proposed, which utilizes the user-generated content, the reply relations among users and the topics of the discussions. The hierarchical model represent the users' interests across different topics. The obtained user profiles are applied to three forum-related tasks: new discussions recommendation, external news articles recommendation and user retrieval. The experimental results show that, comparing with the traditional methods, the hierarchical user profiling approach achieves a better performance in all three tasks.

Getting Clusters from Structure Data and Attribute Data

David Combe, Christine Largeron, Elod Egyed-Zsigmond, Mathias Géry

If the clustering task is widely studied both in graph clustering and in non supervised learning, combined clustering which exploits simultaneously the relationships between the vertices and attributes describing them, is quite new. In this paper, we present different scenarios for this task and, we evaluate their performances and their results on a dataset, with ground truth, built from several sources and containing a scientific social network in which textual data is associated to each vertex and the classes are known. We argue that, depending on the kind of data we have and the type of results we want, the choice of the clustering method is important and we present some concrete examples for underlining this.

Social Network Analysis in Organization Development Studies

Gozde Cüce

This article focuses on the basic concepts, questions and the organization development oriented applications of the Social Network Analysis (SNA) method which has started to be used increasingly by different academicians and practitioners in the fields ranging from medical science to anthropology, from business management to human rights. In the application part of the research, intra-organizational networks of work-related

information, advice, critical support and communication among employees in a bank branch were analyzed and mapped with Social Network Analysis method. By this way, potential fields of intervention in terms of organization development were identified and evaluated.

An Agglomerative Method to Construct Discrepant Cohesive Subgroups

Tobias Hecking, Tilman G☼hnert, H. Ulrich Hoppe

This paper introduces an agglomerative method for detecting cohesive subgroups in networks based on geodesic distance. The algorithm starts with a set of nodes as “seed”. Beginning with the seed nodes as initial clusters, the clusters grow by incorporating more nodes successively based on minimal average distance to the current members of the cluster as a criterion for cluster extension. This approach is combined with an optimization step to achieve high quality performance on subgroup detection. The resulting method for detecting discrepant cohesive subgroups has been tested on artificial benchmark graphs as well as real-world networks.

Naturality of Network Creation Games, Measurement and Analysis

Hedyeh Beyhaghi, Zahra Fahmi, Mohammad Amin Fazli, Jafar Habibi, Pooya Jalaly, Mohammad Ali Safari

Modeling is one of the major research areas in social network analysis whose goal is to study networks structure and its evolution‎. ‎Motivated by the intuition that members in social networks behave selfishly‎, ‎network creation games have been introduced for modeling social networks‎. ‎In this paper‎, ‎our aim is to measure how much the output graphs of a given network creation game are compatible with a social network‎. ‎We first show that the precise measurement is not possible in polynomial time‎. ‎Then we propose a method for its approximation, finally‎, ‎we show the usability of our method by conducting experiments on real network data.

The Limitations of the BP Algorithm for Counting Cycles in Random Networks

Ibrahim Sorkhoh, Khaled Mahdi, Maytham Safar

We study the computation capability of the BP cycles counting algorithm in random networks by estimating the required convergence error to find the maximum number of points in the cycles distribution which helps to interpolate accurately the cycles counts for different sizes. The algorithm shows that the convergence error required to get all possible points is almost constant through all possible connection probability values. The number of points possible can not reach the maximum even with high convergence error values.

Towards Social Version Control

Ali Koc, Abdullah Uz Tansel, Mehmet Bicer

Version control systems are an integral part of software development process since their first use in 1970s. And since then, they have been improved with many features and easy-to-use user interfaces. In this paper, we propose a new feature to be implemented into commonly used version control systems, such as CVS and SubVersion. This feature promotes real time collaboration, opening the door towards a social interaction among development team members. This interaction allows software developers to communicate more effectively through an internal social networking site within the context of the version control environment. The new feature demonstrates that it can be effectively utilized to avoid conflicts, which is one of the major challenges not only in version control, but in any collaborative system.

Community Evolution and Engagement through Assortative Mixing in Online Social Networks

Kon Shing Kenneth Chung, Mahendra Piraveenan, Shahadat Uddin

In this exploratory paper, we examine the evolution and engagement of an online community through a ten-year period. Data is collected from an online public discussion forum provided by a government-sponsored website specifically developed for community capacity building. We postulate that there are clear patterns of assortativity where similar actors engage in communication with each other over time. Results show that there is a clear pattern of networks losing their disassortative character in the early years followed by disassortative networks in the later years. The network-level results challenges government-level metrics of community-building success and suggests network analysis as an empirical avenue for understanding social processes involved in the very nature of community building.

A Navigation Algorithm Inspired by Human Navigation

M. Vijesh, Sudarshan Iyengar, S.M. Vijay Mahantesh, Amitash Ramesh, C. Pandurangan, Veni Madhavan

Human navigation has been a topic of interest in spatial cognition from the past few decades. It has been experimentally observed that humans accomplish the task of way-finding a destination in an unknown environment by recognizing landmarks. Investigations using network analytic techniques reveal that humans, when asked to way-find their destination, learn the top ranked nodes of a network. In this paper we report a study simulating the strategy used by humans to recognize the centers of a network. We show that the paths obtained from our simulation has the same properties as the paths obtained in human based experiment. The simulation thus performed leads to a novel way of path finding in a

network. We discuss the performance of our method and compare it with the existing techniques to find a path between a pair of nodes in a network.

A Probabilistic Inference Attack on Suppressed Social Networks

Baris Altop, Mehmet Ercan Nergiz, Yücel Saygin

Social Networks (SNs) are widely used by internet users to share personal information, which also raises every privacy concern. Hence most service providers offer various preference-based privacy policies, allowing users to suppress any information under their accounts in case they do not wish to share it with public. In this paper, we show that such policies are not sufficient to provide privacy mainly because they do not allow users to control data belonging to other users they are linked with. We show experimentally that one can predict a suppressed boolean label (e.g, being rich or having voted for a specific political party) of a node from other released information in neighboring nodes when there is a known correlation between the links and the label.

Feature Analysis of Spammers in Social Networks with Active Honeypots: A Case Study of Chinese Microblogging Networks

Yi Zhou, Kai Chen, Li Song, Xiaokang Yang, Jianhua He

In this poster we report our study on the microblog spammers with samples attracted by 50 honeyspots from two popular Chinese microblogging networks: Sina Weibo (weibo.com), and Ten cent Weibo (t.QQ.com) in seven months. We studied their features such as social information, activity, account age and spamming strategy. Several distinguishing characteristics of spammers on these two social network communities are observed, which can be helpful to the further study on automatic detection of microblog spammers. To our best knowledge our work is the first of its kind on the analysis of features of Chinese micloblog spammers.

An Inspection Game to Provide Incentive for Cooperation with Corrupted Inspectors

Yalda Kolahdooz, Mohammad Ali Safari

Open and autonomous environments, such as peer to peer networks or many social networks, are efficient only if cooperation among nodes is ensured. In order to ensure cooperative behavior, we have added a new node type to the system, called inspector and used game theoretical tools to analyze the system. Inspectors punish both misbehaving nodes as well as nodes who provide dishonest ratings about other peers. Analyzing the proposed inspection game ensures that corruption of inspectors and misbehavior of nodes is bounded. The game enables the system designer to set the amount of corruption that is allowed according to the budget.

Graph Searching Algorithms for Semantic-Social Recommendation

Dalia Sulieman, Maria Malek, Hubert Kadima, Dominique Laurent

In this paper we present two recommendation algorithms, called Node-Edge-Based and Node-Based recommendation algorithms. These algorithms are designed to recommend items to users connected via social network. Our algorithms are based on three main features: a social network analysis measure (degree centrality), the graph searching algorithm (Depth First Search algorithm), and the semantic similarity measure (which measures the closeness between the input item and users). We apply these algorithms to a real dataset (Amazon dataset) and we compare them with item-based collaborative filtering and hybrid recommendation algorithms. Our results show good precision as well as in a good performance in terms of runtime. Moreover, Node-Edge-Based and Node-Based algorithms search a small part of the dataset, compared to item-based and hybrid recommendation algorithms.

A Web-based Medical Emergency Guiding System

Jui-Hung Kao, Fei-Pei Lai, Wei-Zen Sun, Chia-Ping Shen, Huei-Ming Ma, Jin-Ming Wu, Meng-Yu Chiu, Horng-Twu Liaw, Kai-Chieh Hsu, Yan-Yu Lam, Shih-Ching Cheng

In recent years, the rapid development of trade and traffic in our country has resulted in the increasing occurrence of various types of disasters and emergency injuries. According to the statistics from Fire Departments of Taiwan, medical emergency dispatch frequencies have exceeded 300,000, and the frequency of medical emergency dispatches has increased each year. These figures reflect the phenomenon that, despite the continual growth and progression of the community, various accidents and unforeseen situations are becoming increasingly common. Therefore, fire units have established a medical emergency care system with the objective of minimizing injuries, patient suffering, and the risk of mortality, reducing social medical costs, and alleviating the financial burden of medical emergencies on families. Many countries have special medical emergency dispatch systems, but these do not meet Taiwan's rules and emergency procedures. Thus, our research has focused on compiling a symptom-based medical emergency dispatch guide designed to match Taiwan¡¦s medical emergency dispatch system. We collate national medical emergency dispatch guides and present a medical emergency dispatch system to support medical staff and to educate these staff on the correct methodology in a short time. This system achieves the following objectives: (1) to enhance the quality of medical emergency services in Taiwan, (2) to reduce the response time of medical emergency dispatch, (3) to reduce the unnecessary waste of medical resources.

Temporal Community Structure Patterns in Diabetes Social Networks

Taridzo Chomutare, Eirik Årsand, Gunnar Hartvigsen

The growing amount of data in healthcare social media requires innovative new analysis methods, which are elementary to exploration of relationship dynamics, community formation, sustenance and dissolution, in a bid to understand the new roles social media play in healthcare. In this work we use network analysis to explore the temporal nature of two large diabetes social networks, with a view to enhancing our knowledge of the development of sub-community structures and cohesion factors. Current results reveal how diabetes online communities are very dynamic, suggesting diabetes patients are usually actively engaged for periods of less than a year, typically immediately following diagnosis. The presented empirical evidence inform future online intervention strategies for promoting health behavior and lifestyle changes among people with diabetes.

Developing an Efficient Health Clinical Application: IIOP Distributed Objects Framework

Ayman N. Murshed, Wadhah Almansoori, Konstantinos F. Xylogiannopoulos, Mohamad Elzohbi, Reda Alhajj, Jon Rokne

The Middleware is a piece of software lying between the operating system and the application layer. Distributed applications are gaining popularity with the widespread of reliable communication services. It is affordable to have data accessible 24/7 from almost anywhere, thanks to the well-developed mobile technology and handheld devices. Healthcare domain is one of the very demanding application areas that highly benefits from these developments. Actually, health clinic system is an evolving and promising area in which clients can easily log into a distant health clinic server and retrieve, add, or update the data and diagnosis of the patients using Common Object Request Broker Architecture (CORBA) Internet Inter-Orb Protocol (IIOP) middleware. In this paper, we describe the development and implementation of a middleware model that utilizes the effective connection between two different programming languages (java and .Net) to send and receive requested patients' data in a very efficient response time. This is achieved by using the reference of the object in the server. Based on our knowledge, our approach has the best response time compared to the existing works in the area. Our approach has been successfully tested and evaluated.

A Multiclass Classification Tool Using Cloud Computing Architecture

Chia-Ping Shen, Chia-Hung Liu, Feng-Sheng Lin, Han Lin, Chi-Ying F. Huang, Cheng-Yan Kao, Feipei Lai, Jeng-Wei Lin

Multiclass classification is an important technique to many complex biomedicine problems. Genetic algorithms (GA) are proven to be effective to select features prior to multiclass classification by support vector machines (SVM). However, their use is computation intensive. Based on SOA

(Service Oriented Architecture) design principles, this paper proposes a cloud computing framework that exploits the inherent parallelism of GA-SVM classification to speed up the work. The performance evaluations on an mRNA benchmark cancer dataset have shown the effectiveness and efficiency of the framework. With a user-friendly web interface, the framework provides researchers an easy way to investigate the unrevealed secrets in the fast-growing repository of biomedical data.

Accurate Prediction of Response to Interferon-based Therapy in Egyptian Patients with Chronic Hepatitis C Using Machine-learning Approaches

Mahmoud ElHefnawi, Mahmoud Abdalla, Safaa Ahmed, Wafaa Elakel, Gamal Esmat, Maissa Elraziky, Shaima Khamis, Marwa Hassan

Hepatitis C virus' patients with genotypes 1 & 4 have break-even response rates to Pegylated-Interferon (Peg-IFN) and Ribavirin (RBV) treatment. Furthermore, the incompliance to the treatment because of its high cost and related unfavorable effects makes its prediction of paramount importance. By using machine-learning techniques, a significantly accurate predictive model constructed to predict Egyptian patients' response based on their clinical and biochemical data. The model uses Artificial Neural Networks (ANN) and Decision Trees (DT) to achieve this goal. Two-hundred patients treated with Peg-IFN and RBV, 83 responders (41%), and 117 non-responders (59%) retrospectively studied to extract informative features and train the Neural Networks and Decision Trees. Optimization was done by using six different Neural Network architectures, starting with an input layer of 12 neurons, a hidden layer of 70 to 180 neurons and an output layer containing a single neuron. For decision Trees (DTs), the CART classification algorithm was used. Six DTs with two classes, pruning levels of 9, 11, 13, and 17, and nodes from 45 to 61 were investigated. Among the 12 features in the study, the most statistically significant informative features were the patient's Histology activity index, fibrosis, viral-load, Alfa-feta protein and albumin. Validation of the models on a 20% test set was then performed. The best and average accuracy for the ANN and DT models were 0.76 and 0.69, and 0.80 and 0.72 respectively. Sensitivity and specificity were 0.95 and 0.39, and 0.89 and 0.78 respectively. We conclude that decision trees gave a higher accuracy in predicting response, and would help in proper therapy options for patients.

Improving Tumor Identification by Using Tumor Markers Classification Strategy

Florije Ismaili, Luzana Bekiri

Tumor markers are substances, usually proteins that can be found in the blood, urine, stool, tumor tissue and more recently DNA changes, which are produced by the body in response to cancer growth. Thus far, more than 20 different tumor markers have been identified where some of them are specific for a particular type of cancer, while others are associated with several cancer types. The problem of tumor profiling has been extensively studied by the bioinformatics community. Although tumor classification has improved nowadays, there has been no general

approach for identifying new cancer classes or for assigning tumors to known classes. In this paper we describe a novel strategy for tumor classification by using Growing Hierarchical Self-Organizing map (GHSOM) since it is able to weigh the contribution of each marker according to its relatedness with other tumor markers as well as handles highly skewed tumor marker expressions well. In the end, experiments are conducted to further demonstrate the feasibility and efficiency of tumor classification approach which provide valuable contribution in the field of oncology and cancer diseases and will be as a guide for the identification of these diseases.

A Classifier to Detect Tumor Disease in MRI Brain Images

Amer Al-Badarneh, Hassan Najadat, Ali M. Alraziqi

The traditional method for detecting the tumor diseases in the human MRI brain images is done manually by physicians. Automatic classification of tumors of MRI images requires high accuracy, since the non-accurate diagnosis and postponing delivery of the precise diagnosis would lead to increase the prevalence of more serious diseases. To avoid that, an automatic classification system is proposed for tumor classification of MRI images. This work shows the effect of neural network (NN) and K-Nearest Neighbor (K-NN) algorithms on tumor classification. We used a benchmark dataset MRI brain images. The experimental results show that our approach achieves 100% classification accuracy using K-NN and 98.92% using NN.

Comparing Clustering Techniques for Real Microarray Data

Vilda Purutçuoglu Gazi, Elif Kayis

The clustering of genes detected as significant or differentially expressed provides useful information to biologists about functions and functional relationship of genes. There are variant types of clustering methods that can be applied in genomic data. These are mainly divided into the two groups, namely, hierarchical and partitional methods. In this paper, as the novelty, we perform a detailed clustering analysis for the recently collected boron micro array dataset to investigate biologically more interesting results and to construct a basis for the selection of the most effective method in the analysis of different micro array datum. In the calculation, we implement the agglomerative hierarchical clustering among hierarchical techniques and use the k-means and the PAMSAM methods within partitional clustering approaches, and finally use a recently improved method, called HIPAM, which is not only a hierarchical but also partitional approach. In the assessment, we compare and discuss the significant genes of the boron data whose estimated signals are found by the FGX normalization method.

Automatic Drug Adverse Reaction Discovery from Parenting Websites Using Disproportionality Methods

Jelena Hadzi-Puric, Jeca Grmusa

The pediatric fact that children "are not just little adults" reflects special drug class, dosing, pharmacokinetic parameters that must be considered when planning safe and effective health interventions, because evidence of drug safety and efficiency cannot be extrapolated from adults to children. There is significant interest in utilization of observational health care data in pediatric drug safety research. We gathered data from eight different parenting websites in order to explore automatic drug adverse reaction discovery. In recent years, Web forums offer an opportunity for parents and doctors to share healthcare information regarding prescription drugs. Our aim is to discover drugs and adverse reactions and quantify their association using statistical methods based on different measures of disproportionality. Also, applied disproportionality methods have the potential to scale well to large databases such ours.

Newborn Screening for Phenylketonuria: Machine Learning vs Clinicians

Wei-Hsin Chen, Han-Ping Chen, Yi-Ju Tseng, Kai-Ping Hsu, Sheau-Ling Hsieh, Yin-Hsiu Chien, Wuh-Liang Hwu, Feipei Lai

The metabolic disorders may hinder an infant¡¦s normal physical or mental development during the neonatal period. The metabolic diseases can be treated by effective therapies if the diseases are discovered in the early stages. Therefore, newborn screening program is essential to prevent neonatal from these damages. In the paper, a support vector machine (SVM) based algorithm is introduced in place of cut-off value decision to evaluate the analyte elevation raw data associated with Phenylketonuria. The data were obtained from tandem mass spectrometry (MS/MS) for newborns. In addition, a combined feature selection mechanism is proposed to compare with the cut-off scheme. By adapting the mechanism, the number of suspected cases is reduced substantially, it also handles the medical resources effectively and efficiently.

Wavelet-based Multiscale Filtering of Genomic Data

Mohamed Nounou, Hazem Nounou, Nader Meskin, Aniruddha Datta

Measured biological data are a rich source of information about the biological phenomena they represent. For example, time-series genomic or metabolic micro array data can be used to construct dynamic genetic regulatory network models, which can be used to better understand the biological system and to design intervention strategies to cure or manage major diseases. Unfortunately, biological measurements are usually highly contaminated with errors that mask the important features in the data. Therefore, these noisy measurements need to be filtered to

enhance their usefulness in practice. Wavelet-based multiscale filtering has been shown to be a powerful data analysis and denoising tool. In this work, different batch as well as online multiscale filtering techniques are used to filter biological data contaminated with white noise. The performances of these multiscale filtering techniques are demonstrated and compared to those of some conventional low pass filters using simulated time series metabolic data. The results of this comparative study show that significant improvement can be achieved using multiscale filtering over conventional filtering methods.

Robot-Assisted Medical Visualization with Floating Images

Sandor Markon, Satoshi Maekawa, Ahmet Onat

Accessing volumetric medical data requires advanced visualization techniques. One approach is to show slices in situ, by projecting images into space at the proper position. We propose using floating images, that is, undistorted real images of displays, appearing at the desired position and orientation. Such floating images are made possible by the new optical device eDCRAf invented at NICT in Japan. To enhance the freedom of the user, we have developed a robot-assisted interactive visualization system, where the user can eholdf the image slice in her/his hand, and freely change its position and orientation, thus inspect any part of the volumetric data set as desired. The hand position is detected by a stereo camera and image processing. We describe the structure and operation of our system, and show samples of its usage.

Security Standards for Electronic Health Records

Oznur Esra Par, Ergin Soysal

Circulation of personal health records in digital media has increased by intensive usage of technology on health sector. Circulation of personal health records in electronic media brings with security and privacy issues. Electronic health records are all of this information which includes patient data from birth to the death of the patient. Since electronic health records include private and unchangeable information, it is being tried to forbid their usage revelation without permission in accordance with the related legislations. Digitization of personal health records also brings with security risks. A number of technical and legal infrastructure is needed to eliminate these risks. With the scope of the research national (such as HIPAA) and international standards (such as ISO) has been studied.

Response Surface Modeling and Optimization to Elucidate the Differential Effects of Demographic Characteristics on HIV Prevalence in South Africa

Wilbert Sibanda, Philip Pretorius, Anne Grobler

In this study, a Central Composite Face Centered (CCF) design was employed to study the individual and interaction effects of demographic characteristics on the spread of HIV in South Africa. The demographic characteristics studied for each pregnant mother attending an antenatal clinic in South Africa, were mother's age, partner's age, mother's level of education and parity. HIV status of an antenatal clinic attendee was found to be highly sensitive to changes in pregnant woman's age and partner's age, using the 2007 South African annual antenatal HIV and syphilis seroprevalence data. Individually the pregnant woman's level of education and parity had no significant effect on the HIV status. However, the latter two demographic characteristics exhibited significant effects on the HIV status of antenatal clinic attendees in two way interactions with other demographic characteristics. Using HIV as the optimization objective, the following summary statistics were obtained, R^2 = 0.99 and two-factor interactions (2FI) model F-value of 63.77. The model F-value of 63.77 implied the 2FI model was significant and there was only a 0.01% chance this model value could occur due to noise. The model 'Lack of Fit' value of 0.01 implied that the 'Lack of Fit' was not significant relative to the pure error and thus there was a 99.88% chance that this 'Lack of Fit' F-value could occur due to noise. An adeq. precision value of 25 was obtained, suggesting that this 2FI model could be used to navigate the design space. A 3D response surface plot indicated that the highest rate of HIV positive individuals was obtainable at the highest age of the pregnant women and lowest age of their partners.

A Novel Framework for Spammer Detection in Social Bookmarking Systems

Soghra M.Gargari, Sule Gündüz Ogüdücü

Content Mining of Microblogs

M. Ozgür Cingiz, Banu Diri

Emergence of Web 2.0, internet users can share their contents with other users using social networks. In this paper microbloggers' contents are evaluated with respect to how they reflect their categories. Migrobloggers' category information, which is one of the four categories that are economy sport, entertainment or technology, is taken from wefollow.com application. 2105 RSS news feeds, whose category labels are same with microbloggers' contributions, are used as training data for classification. In this study two types of users' contributions are taken as test data. These users are normal micro loggers and bots. Classification results show that bots provide more categorical content than normal users.

An Application Based on Steganography

H.B. Karaman, S. Sagiroglu

In this study, the aim and application areas of steganography and steganalysis are briefly described. Kinds of techniques used in steganography are analyzed and compared to each other according to their simplicity and robustness. Then, methods used in steganalysis in order to detect various kinds of steganography such as text, image, audio, video and file system steganography are examined. Finally, a steganography application for the receptions of the Presidency of the Republic of Turkey is developed. The identity information of the guests are first encrypted, and then embedded in their photos using the LSB technique on BMP images. Original and stego images after the embedding process are compared to each other and the robustness of the application against steganalysis attacks is measured using VSL tool.

A Model for Dynamic Integration of Data Sources

Murat Obali, Bunyamin Dursun

Online and offline data is the key to Intelligence Agents, but these data cannot be fully analyzed due to the wealth and complexity and non-integrated nature of the information available. In the field of security and intelligence, there is a huge number of data coming from heterogonous data sources in different formats. The integration and the management of these data are very costly and time consuming. The result is a great need for dynamic integration of these intelligent data. In this paper, we propose a complete model that integrates different online and offline data sources. This model takes part between the data sources and our applications.

Combining Entity Matching Techniques for Detecting Extremist Behavior on Discussion Boards

Johan Dahlin, Fredrik Johansson, Lisa Kaati, Christian M▫rtenson, Pontus Svenson

Many extremist groups and terrorists use the Web for various purposes such as exchanging and reinforcing their beliefs, making monitoring and analysis of discussion boards an important task for intelligence analysts in order to detect individuals that might pose a threat towards society. In this work we focus on how to automatically analyze discussion boards in an effective manner. More specifically, we propose a method for fusing several alias (entity) matching techniques that can be used to identify authors with multiple aliases. This is one part of a larger system, where the aim is to provide the analyst with a list of potential extremist worth investigating further.

Critical Infrastructure and Internal Controls

Iztok Podbregar, Mojca Ferjancic Podbregar

Critical refers to infrastructure that provides an essential support for economic and social well-being, for public safety and for the functioning of key government responsibilities. According to Resolution of the National Security Strategy of the Republic of Slovenia various sources of threat are directed towards critical infrastructure. An operational infrastructure ensures the implementation of key state and society function. In this respect, the treat to critical infrastructure may also influence the national security. Risk is the possibility of an event occurring that will have an impact on the achievement of objectives. When we discuses risks, we usually talk about negative impacts, which can cause that we can't be able to achieve our objectives. Risk management is an act or process of controlling risks. All of the government programmes studied adopt a risk management approach to critical infrastructure protection. Risk management helps governments to identify key security assets, assess risks and establishes strategies and priorities for mitigating these risks. We can almost never eliminate the risks completely. Risks can be reduced to an acceptable level. The acceptable risk is a set of risks that the process is facing is sufficiently small or unimportant for the critical infrastructure to coexist with it and not try to further reduce it. Management or those who are responsible for critical infrastructure will want to reduce risk to an acceptable level by applying controls and other measures.

Covertness Centrality in Networks

Michael Ovelgⵔnne, Chanhyun Kang, Anshul Sawant, V.S. Subrahmanian

It has been known for some time that in terror networks, money laundering networks, and criminal networks, "important" players want to stay "off" the radar. They need sufficient centrality (according to traditional measures) to be well connected with the rest of their network, but need to blend in with the crowd. In this paper, we propose the concept of covertness centrality (CC). The covertness centrality of a vertex v consists of two parts: how "common" v is w.r.t. a set $\mathcal{C}$ of centrality measures, and how well v can "communicate" with a user-specified set of vertices. The more "common" v is, the more able it is to stay hidden in a crowd. Given $\mathcal{C}$, we first propose some general properties we would like a common-ness measure to satisfy. We then develop a probabilistic model of common-ness that a vertex has w.r.t. $\mathcal{C}$ (specifying, intuitively, how many other vertices are like it according to all centrality measures in $\mathcal{C}$). Covertness centrality of vertex v is then defined as a linear combination of common-ness and the ability of v to communicate with a user-specified set of other vertices. We develop a prototype implementation of CC and report on experiments we have conducted with it on several real-world data sets.

Combining Spatial Proximity and Temporal Continuity for Learning Invariant Representations

Olcay Kursun, Tevfik Aytekin

Location and time are two critical aspects of most security-related events, and thus, spatiotemporal data analysis plays a central role in many security-related applications. The human brain has great capabilities of developing invariant representations of objects by taking advantage of both spatial similarity of features of objects/events and their relative timings (temporal information). Trace learning rule is one well-known solution for this problem of combining temporal relations with spatial proximity in clustering tasks such as the one performed by self organizing maps. In this work, we investigate a two stage mechanism: i) finding local clusters using spatial proximity, ii) grouping these clusters as suggested by temporal continuity patterns. We show our experimental results on a movie created from face images.

Mining Divergent Opinion Trust Networks through Latent Dirichlet Allocation

Nima Dokoohaki, Mihhail Matskin

While the focus of trust research has been mainly on defining and modeling various notions of social trust, less attention has been given to modeling opinion trust. When speaking of social trust mainly homophily (similarity) has been the most successful metric for learning trustworthy links, specially in social web applications such as collaborative filtering recommendation systems. While pure homophily such as Pearson coefficient correlation and its variations, have been favorable to finding taste distances between individuals based on their rated items, they are not necessarily useful in finding opinion distances between individuals discussing a trending topic, e.g. Arab spring. At the same time text mining techniques, such as vector-based techniques, are not capable of capturing important factors such as saliency or polarity which are possible with topical models for detecting, analyzing and suggesting aspects of people mentioning those tags or topics. Thus, in this paper we are proposing to model opinion distances using probabilistic information divergence as a metric for measuring the distances between people's opinion contributing to a discussion in a social network. To acquire feature sets from topics discussed in a discussion we use a very successful topic modeling technique, namely Latent Dirichlet Allocation (LDA). We use the distributions resulting to model topics for generating social networks of group and individual users. Using a Twitter dataset we show that learned graphs exhibit properties of real-world like networks.

Mining Hate Crimes to Figure Out Reasons Behind

Fatih Ozgul, Murat Gok, Yakup Ozal, Ahmet Celik

In order to identify the reasons behind the hate crimes in Diyarbakir, Turkey a data mining research has been made. Various data mining algorithms are applied to data set containing forty big cases happened between 2009 and 2010. some algorithms helped us to model criminality in Diyarbakir, we learned which features for hatred crime features are important while dealing with cases. There has been also contribution for identifying habits of Diyarbakir people which makes them vulnerable for the hate crimes.

The Mental State of Influencers

D.B. Skillicorn, C. Leuprecht

Most analysis of influence looks at the mechanisms Used, and how effectively they work on the intended audience. Here we consider influence from another perspective: what do the language choices made by influencers enable us to detect about their internal mental state, strategies and assessments of success. We do this by examining the language used by the U.S. presidential candidates in the high-stakes attempt to get elected. Such candidates try to influence potential voters, but must also pay attention to the parallel attempts by their competitors to influence the same pool. We examine seven channels: persona deception, the attempt by each candidate to seem as attractive as possible, nouns, as surrogates for content; positive and negative language; and three categories that have received little attention, verbs, adverbs, and adjectives. Although the results are preliminary, several intuitive and expected hypotheses are supported, but some unexpected and surprising structures also emerge. The results provide insights into related influence scenarios where open-source data is available, for example marketing, business reporting, and intelligence.

Perspective Analysis for Online Debates

Sukru Tikves, Sedat Gokalp, Mhamed Temkit, Sujogya Banerjee, Jieping Ye, Hasan Davulcu

Internet and social media devices created a new public space for online debate on political and social topics. A debate is defined as a formal discussion on a set of related topics in a public meeting, in which opposing perspectives and arguments are put forward. In this paper, we develop automated perspective discovery techniques which would contribute to the understanding of features (i.e. social, political, cultural, religious beliefs, goals, and practices) shared by each side of the debate. Secondly, we show that, compared to a semi-automated process, our

perspective discovery algorithms not only identify larger number of relevant features, but they also yield a higher accuracy scaling of moderate to extreme organizations on both sides of a debate.

Processes View Modeling of Identity-related Privacy Business Interoperability: Considering User-Supremacy Federated Identity Technical Model and Identity Contract Negotiation

Ghazi Ben Ayed, Solange Ghernaouti-Hélie

Federated identity is a distributed system that is deployed across multiple parties. Service providers still hold the absolute power over people identities. So, identity-related privacy is considered as a mean to entrench subjects' control over identities and foster trust among multiple involved parties. Thus, identity-related privacy should interoperable, which can be guaranteed through the capture of requirements from different polices related to identity. In this article, we provide and explain a BPMN processes view of the requirements allowing them to be ready to-implement, clear, easy to-understand by each party wishing to collaborate within or across federated identity systems. We highlight that present-day practitioners should be able to translate requirements with user-supremacy federated identity technical model concepts into a set of rules and take into consideration details of identity contract negotiation in order to successfully deliver processes view. BPMN collaboration and choreography diagrams are used to describe seven processes and a sub-process, which would provide a useful way to gain alignment between requirements and IT.

Secret Sharing Scheme: Vector Space Secret Sharing and <i>Φ</i> Function

Mustafa Atici

Let ${\cal P}=\{P_{1}, P_{2}, ..., P_{n}\}$ be set of participants and $\Gamma=\{B_{i}| B_{i}\subset {\cal P}, 1\leq i \leq k \}$ be access structure. Vector space secret sharing realizing access structure Γ requires existence of function $\phi : {\cal P}\longrightarrow ({\cal Z}_{p})^{d}$, where p is a prime number and $d\geq 2$ is an integer, satisfying the following condition $(1, 0, 0, ..., 0)= \Leftrightarrow B \in \Gamma=\{ B_{1}, B_{2}, ..., B_{k}\}$. There is no known algorithm to construct such a function ϕ in general. Constructions are mainly done by trial and error. In this paper, we developed a polynomial algorithm to construct a ϕ function for certain type of access structures. Some examples are given to illustrate the algorithms.

Sentiment Analysis on Social Media

Federico Neri, Carlo Aliprandi, Federico Capeci, Montserrat Cuadros, Tomas By

The Web is a huge virtual space where to express and share individual opinions, influencing any aspect of life, with implications for marketing and communication alike. Social Media are influencing consumersâ€™ preferences by shaping their attitudes and behaviors. Monitoring the Social Media activities is a good way to measure customersâ€™ loyalty, keeping a track on their sentiment towards brands or products. Social Media are the next logical marketing arena. Currently, Facebook dominates the digital marketing space, followed closely by Twitter. This paper describes a Sentiment Analysis study performed on over than 1000 Facebook posts about newscasts, comparing the sentiment for Rai - the Italian public broadcasting service - towards the emerging and more dynamic private company La7. This study maps study results with observations made by the Osservatorio di Pavia, which is an Italian institute of research specialised in media analysis at theoretical and empirical level, engaged in the analysis of political communication in the mass media. This study takes also in account the data provided by Auditel regarding newscast audience, correlating the analysis of Social Media, of Facebook in particular, with measurable data, availbale to public domain.

Uncovering Mobile Phone Users' Malicious Activities Using Open Source Tools

P.A. Kotsopoulos, Y.C. Stamatiou

Smart mobile phones are, nowadays, in wide use and day by day they are transformed into an essential part of everyday life, for business and for entertainment alike. Mobile Networks have been implemented in order to serve these needs. Unfortunately there are many users who want to use the benefits that technology gives to them for malicious and/or illegal actions. In this paper we present and analyze a hypothetical (but realistic) cyber crime investigation scenario. According to this, a user who has wireless internet access using a mobile phone through a University's network facilities, visits sites with pornographic content. A fact that complicates the analysis of the scenario is that the user employs anti-forensics techniques (e.g. data encryption). In this paper we show how to apply to the investigation of this scenario (as well as similar ones) a number of widely available open source forensics analysis tools overcoming the anti-forensics efforts of malicious users. Our goal is to demonstrate their applicability to cyber-crime forensics and their capability to operate in a complementary, to each other, way so that their deployment sheds light to different aspects of the case at hand.

Tag Based Recommender System for Social Bookmarking Sites

Fatemeh Ghiyafeh Davoodi, Omid Fatemi

It is often essential for people to consult with others and ask them about their past experience and thoughts when making choices. Exchanging ideas among people has become more meaningful since the extensive growth of information on the World Wide Web (WWW). People have access to tremendous amount of information, but choosing the most relevant information is of high effort. It was when recommender systems came into existence in 1992 in order to assist users in the process of finding the most appropriate information on WWW, and identify sets of items which are likely to be interesting for the users. Recommender systems have used different sources of data in order to identify users' interests. In addition, by growth of social resource sharing like social book marking sites, tagging activities can be considered as explicit knowledge for user and item modeling. Existing recommender systems lack use of external source of information for recommending the most appropriate item. They mainly use the information of their own website, while there is valuable information on the web which could improve the performance of the predictions. In this paper, we use Open Directory Project (ODP) data as external knowledge about web pages in addition to tagging activities of users in a social book marking site. We have designed a content based recommender system which can recommend the most relevant web pages for each user based on the user's profile and gathered information about web pages from ODP as implicit data. We empirically evaluate effect of ODP data on the predictions using Delicious dataset in order to analyze the performance of the proposed method. The results show that our recommender system outperforms when it uses ODP information as external source of data.

A Hybrid Approach for Biometric Template Security

Kareem Kamal A. Ghany, Hesham A. Hefny, Aboul Ella Hassanien, Neveen I. Ghali

Privacy and Security has become an increasingly serious ýproblem for any biometric systems. Template protection, ýwhich mainly prevents from data loss and hacking the stored ýtemplates, is one of the most important issues when considering ýprivacy and security. Cancelable biometrics approach Scheme ýhas been proposed to address this problem. On the other hand ýthe symmetric hash functions might be used to increase the ýbiometric security level by making it is hard to attack the ýtemplate. In this paper, a new hybrid approach based on ýcombining approaches based on transformation and ýcryptosystem on fingerprints is applied. The results are ýcompared to a pre-presented approach in [1] showing that the ýnew approach is more efficient as the security of biometric ýtemplate is increased meanwhile the error rate is minimized.

Analyzing Change over Time in Organizations' Publics with a Semantic Network Include List: An Illustration with Facebook

James A. Danowski

This research highlights a kind of semantic network analysis based on an include list. We analyze the networks only among words on the list as they appear in a series of text corpora word pairs for an organization. The example uses documents about Facebook over a 12-month period, dividing them into 12 time-based files. In each time slice we map networks among key publics and measure the centrality of each from one time period to the next. The network of publics becomes more complex across time. Publics fluctuate in centrality. We describe other kinds of semantic network analysis for business applications using include lists.

Women on Public-Company Boards: Factors That Affect their Odds of Board Membership Relative to Those of Men

Renuka Hodigere, Diana Bilimoria

Despite anecdotal support, empirical evidence for network effects on women's advancement to corporate board membership has yet to be fully documented. To address these gaps in extant knowledge, the primary questions that the proposed project seeks to answer how network variables affect the likelihood of board membership for women differently from men? We found, from examining the networks of 489 directors, constraints in the opportunity structure available to women in the upper echelons of organizations.

Sharing News Articles Using 140 Characters: A Diffusion Analysis on Twitter

Devipsita Bhattacharya, Sudha Ram

Is it possible to effectively spread news articles to a large audience using 140 characters? How does the microblogging website Twitter get used as a platform for the news media agencies to create awareness about the articles they publish on a daily basis? Our study of the diffusion patterns of news articles from 12 popular news sources, including BBC, New York Times, and Mash able on Twitter reveals that a large number of users not only consume and comment on these news articles but also share them in different ways. Combining the methods of network and temporal analyses, we examine and report on how news articles diffuse on Twitter, and how different propagation mechanisms result in different life spans for news articles.

Monitoring User Evolution in Twitter

Claudia Lauschke, Eirini Ntoutsi

Nowadays social media are widely used for the broadcasting of different types of information, such as events, activities and opinions. Analyzing this vast amount of data for extracting models that describe individual users or groups of users has gained a lot of attention lately. In this work we analyze individual users and monitor changes in their published content over time. We propose a topic-based user profiling and monitoring approach for change detection and monitoring of profile evolution. Our method is capable of detecting persistent topics representing long term interests of the user as well as short term topics that refer to everyday events or reactions to the news. We evaluate our approach on real data from Twitter.

Spatio-Temporal Web Sensors by Social Network Analysis

Shun Hattori

Many researches on mining the Web, especially Social Networking Media such as web logs and microblogging sites which seem to store vast amounts of information about human societies, for knowledge about various phenomena and events in the physical world have been done actively, and Web applications with Web-mined knowledge have begun to be developed for the public. However, there is no detailed investigation on how accurately Web-mined data reflect real-world data. It must be problematic to idolatrously utilize the Web-mined data in public Web applications without ensuring their accuracy sufficiently. Therefore, this paper defines spatio-temporal Web Sensors by analyzing Twitter, Facebook, web logs, news sites, or the whole Web for a target natural phenomenon, and tries to validate the potential and reliability of the Web Sensors' spatio-temporal data by measuring the coefficient correlation with Japanese weather, earthquake, and influenza statistics per week by region as real-world data.

A Probabilistic Approach to Structural Change Prediction in Evolving Social Networks

Krzysztof Juszczyszyn, Adam Gonczarek, Jakub M. Tomczak, Katarzyna Musial, Marcin Budka

We propose a predictive model of structural changes in elementary sub graphs of social network based on Mixture of Markov Chains. The model is trained and verified on a dataset from a large corporate social network analyzed in short, one day-long time windows, and reveals distinctive patterns of evolution of connections on the level of local network topology. We argue that the network investigated in such short timescales is highly dynamic and therefore immune to classic methods of link prediction and structural analysis, and show that in the case of complex

networks, the dynamic sub graph mining may lead to better prediction accuracy. The experiments were carried out on the logs from the Wroclaw University of Technology mail server.

A Repository for Multirelational Dynamic Networks

Alexander Semenov, Jari Veijalainen

Nowadays, WWW contains a number of social media sites, which are growing rapidly. One of the main features of social media sites is to allow to its users creation and modification of contents of the site utilizing the offered WWW interfaces. Such contents are referred to as user generated contents and their type varies from site to site. Social media sites can be modeled as constantly evolving multirelational directed graphs. In this paper we discuss persistent data structures for such graphs, and present and analyze queries performed against the structures. We also estimate the space requirements of the proposed data structures, and compare them with the naive “store each complete snapshot of the graph separately”. We also investigate query performance against our data structure. We present analytical estimation results, simulation results, and discuss its performance when it is used to store entire contents of Live journal.

Capturing Actor-level Dynamics of Longitudinal Networks

Shahadat Uddin, Kon Shing Kenneth Chung, Mahendra Piraveenan

Study of the dynamics of longitudinal networks has already attracted enormous research interest. Although dynamics of networks can be captured both at network-level and node / actor-level, the latter has gained less attention in current literature. By following a topological approach (i.e., static topology and dynamic topology) to analyze networks, this paper first proposes a research framework to capture actor-level dynamics for longitudinal networks. In static topology, Social Network Analysis (SNA) methods are applied on the aggregated network of entire data collection period. A smaller segment of network data that are accumulated in less time compared to the entire data collection period are used in dynamic typology for analysis purpose. This study further successfully compiles and applies this framework to the context of organizational crisis and project dynamics with the purpose to explore different level of actor-level dynamics at the different operational environment of these contexts over time. It is noticed that different level of actor-level dynamics are observed in the communication and collaboration network during the different facets of the organizations. In the context of organizational crisis, it is evident that during the ‘crisis' period of operational running of organization, actors in the organizational email communication networks show higher level of actor-level dynamics compared to the ‘normal' period. Less actor-level dynamics are observed during the ‘final' phase of project life cycle, as found from the second context.

Social Influence Estimation for Short Texts in Plurk

Han-Chih Liu, Jenq-Haur Wang

Social media present a user-friendly way of communication and sharing, which brings new chances to understand users and their social communication patterns. With the popularity of microblogging services, the huge volume of very short texts makes it difficult to track the latest updates or breaking news. In this paper, we propose a novel social influence model for estimating the popularity score for each short text in plurk. First, the degrees of user participation and user propagation are estimated by the number of replies, replurks, likes, and URIs. Then, we measure the influence persistence by the duration of the initial post and the last response, and the influence score can be derived from a linear combination of these simple statistics. Our experimental results on more than 300 thousand plurks collected from 1,750 users showed a good performance in determining popular messages, with the best F-measure of 0.86. From our case studies, top-ranked messages can accurately reflect the popular discussions on important events. This shows the effectiveness of our proposed approach. Further investigation of applying the influence model in event detection is needed.

A Social Network Analysis of Interlocking Directorates in French Firms

Sana Elouaer-Mrizak

Using Social Network Analysis, this study investigates whether a board interlocking network exists among the French listed companies (we consider the French Financial Indices CAC 40 and SBF 250) and how this network differs between CAC40 companies and SBF250 companies. Moreover, this research focuses on whether a firmâ€™s centrality is associated with its size. The results suggest that (1) the interlocking network is denser in 1996 and 2010 than in 2005, (2) big companies (size is measured by the firmâ€™s market capitalization) are the central nodes in these networks, (3) the location of the company is more important than simply the number of ties.

T-Recs: Time-aware Twitter-based Drug Recommender System

Ahmed Abdeen Hamed, Rebecca Roose, Marlon Branicki, Alan Rubin

We introduce our initial work for developing a social networks recommender system called T-Recs. The system is a time-aware Twitter-based alternative medicine recommender system. We collected a set of tweets (around 500,000 tweets) that contain specific hash tags (#throwing up, #headache, #itching etc) for a three consecutive weeks. The individual tweets were examined manually by a domain expert (a medical doctor) who inspected the tweet sentiments along with the tweet's timestamp. Using this data, the domain expert assigned a preliminary label to the

tweet/group of tweets. We then trained a classifier using the hash tags and its labels taking into consideration other factors (i.e., age, gender, co-morbidity conditions) that the Tweeter must provide by taking a questionnaire. After the questions are answered, the recommender system makes recommendations based on the Tweet contents and the questionnaire factors for never-seen-before tweets. The classifier, is the core component of the recommender system, is designed as a Decision Tree algorithm to classify the Tweets. Each recommendation is made by the system provides a medical advice to promote public health awareness, and a link to the recommender systems' web portal to answer the questionnaire. When the questionnaire is submitted, the symptoms and Tweeter's basic information are consolidated and a recommendation is made available to the user at once. For the sake of feasibility, we only considered tweets that are sent from within the United States.

Classification Analysis in Complex Online Social Networks Using Semantic Web Technologies

Marek Opuszko, Johannes Ruhland

The Semantic Web enables people and computers to interact and exchange information. Based on Semantic Web technologies, different machine learning applications have been designed. Particularly important is the possibility to create complex metadata descriptions for any problem domain, based on pre-defined ontologies. In this paper we evaluate the use of a semantic similarity measure based on pre-defined ontologies as an input for a classification analysis in the context of social network analysis. A link prediction between actors of two real world social networks is performed, which could serve as a recommendation system. The social networks involve different types of relations and nodes. We measure the prediction performance based on a semantic similarity measure as well as traditional approaches. The findings demonstrate that the prediction accuracy based on the semantic similarity is comparable to traditional approaches and shows that data mining on complex social networks using ontology-based metadata can be considered as a very promising approach.

How Well Do We Know Each Other?” Detecting Tie Strength in Multidimensional Social Networks

Luca Pappalardo, Giulio Rossetti, Dino Pedreschi

The advent of social media have allowed us to build massive networks of weak ties: acquaintances and nonintimate ties we use all the time to spread information and thoughts. Conversely, strong ties are the people we really trust, people whose social circles tightly overlap with our own and, often, they are also the people most like us. Unfortunately, the majority of social media do not incorporate explicitly tie strength information in the creation and management of relationships, and treat all users the same: friend or stranger, with little or nothing in between. In the current work, we address the challenging issue of detecting on online social networks the strong and intimate ties from the huge mass of such mere social contacts. In order to do so, we propose a novel multidimensional definition of tie strength which exploits the existence of

multiple online social links between two individuals. We test our definition on a multidimensional network constructed over users in Foursquare, Twitter and Facebook, analyzing the structural role of strong and weak links, and the correlations with the most common similarity measures.

Distributed Assessment of Network Centralities in Complex Social Networks

Klaus Wehmuth, Artur Ziviani

We present our ongoing work towards a framework for the Distributed Assessment of Network Centralities (DANCE) in complex networks. DANCE targets the efficient evaluation of the network centrality based only on localized information, restricted to a limited neighborhood around each node of the network. In this way, it does not require full knowledge of the network topology. We also discuss how DANCE can be tailored to evaluate different kinds of network centrality. This outcome is quite useful given the vast potential applicability of centrality-based analysis on the large-scale complex social networks.

Cyclic Entropy of Complex Networks

Ibrahim Sorkhoh, Khaled Mahdi, Maytham Safar

We calculate the cyclic entropy of a real virtual friendship network to have an insight on the degree of its robustness. Upon counting the number of cycles of different sizes in the network, a probability distribution function is resulted. An actual friendship network is found to have cyclic entropy bounded between random and small-world networks models. It has dual properties. Small world networks indicate the existence of critical network sizes: 150 and 700 at which the cyclic entropy is minimum. Scale-free networks have the highest cyclic entropy among all other complex network models regardless of the size of the network.

Predicting Social Network Measures Using Machine Learning Approach

Radoslaw Michalski, Przemyslaw Kazienko, Dawid Król

The link prediction problem in social networks defined as a task to predict whether a link between two particular nodes will appear in the future is still a broadly researched topic in the field of social network analysis. However, another relevant problem is solved in the paper instead of individual link forecasting: prediction of key network measures values, what is a more time saving approach. Two machine learning techniques were examined: time series forecasting and classification. Both of them were tested on two real-life social network datasets.

Studying User Footprints in Different Online Social Networks

Anshu Malhotra, Luam Totti, Wagner Meira_Jr., Ponnurangam Kumaraguru, Virgﻱlio Almeida

With the growing popularity and usage of online social media services, people now have accounts (some times several) on multiple and diverse services like Facebook, Linked In, Twitter and You Tube. Publicly available information can be used to create a digital footprint of any user using these social media services. Generating such digital footprints can be very useful for personalization, profile management, detecting malicious behavior of users. A very important application of analyzing users' online digital footprints is to protect users from potential privacy and security risks arising from the huge publicly available user information. We extracted information about user identities on different social networks through Social Graph API, Friend Feed, and Profilactic, we collated our own dataset to create the digital footprints of the users. We used username, display name, description, location, profile image, and number of connections to generate the digital footprints of the user. We applied context specific techniques (e.g. Jaro Winkler similarity, Word net based ontologies) to measure the similarity of the user profiles on different social networks. We specifically focused on Twitter and Linked In. In this paper, we present the analysis and results from applying automated classifiers for disambiguating profiles belonging to the same user from different social networks. User ID and Name were found to be the most discriminative features for disambiguating user profiles. Using the most promising set of features and similarity metrics, we achieved accuracy, precision and recall of 98%, 99%, and 96%, respectively.

FakeBook: Detecting Fake Profiles in On-Line Social Networks

Mauro Conti, Radha Poovendran, Marco Secchiero

On-line Social Networks (OSNs) are increasingly influencing the way people communicate with each other and share personal, professional and political information. Like the cyberspace in Internet, the OSNs are attracting the interest of the malicious entities that are trying to exploit the vulnerabilities and weaknesses of the OSNs. Increasing reports of the security and privacy threats in the OSNs is attracting security researchers trying to detect and mitigate threats to individual users. With many OSNs having tens or hundreds of million users collectively generating billions of personal data content that can be exploited, detecting and preventing attacks on individual user privacy is a major challenge. Most of the current research has focused on protecting the privacy of an existing online profile in a given OSN. Instead, we note that there is a risk of not having a profile in the last fancy social network! The risk is due to the fact that an adversary may create a fake profile to impersonate a real person on the OSN. The fake profile could be exploited to build online relationship with the friends of victim of identity theft, with the final target of stealing personal information of the victim, via interacting online with the friends of the victim. In this paper, we report on the investigation we did on a possible approach to mitigate this problem. In doing so, we also note that we are the first ones to analyze social network graphs from a dynamic point of view within the context of privacy threats.

An Improved Model of Trust-aware Recommender Systems Using Distrust Metric

Ali Nazemian, Hoda Gholami, Fattaneh Taghiyareh

Trust–aware recommender systems are intelligent technology applications that make use of trust information and user personal data in social networks to provide personalized recommendations. Recent research on recommender systems shows that these recommender systems are more robust against shilling attacks and can better be used for generating recommendations for new users. In this paper we proposed a model for improving the accuracy of trust-aware recommender systems. The results of evaluating our approach on Extended Epinions dataset shows that this approach can improve accuracy of recommender systems significantly while does not reduce the coverage of recommender systems.

Optimal Clustering Selection on Hierarchical System Network

Eddie Fuller, Wenliang Tang, Yezhou Wu, Cun-Quan Zhang

In data mining, hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types: agglomerative and divisive. In this paper we shall introduce a new optimal selection method based on the well-known Max-Flow Min-Cut theorem, which also works for the hierarchically structure with overlapping. A novel dynamic algorithm was presented for the special structure without overlapping.

BidTerm Suggestion for Advertising Webpages

Lamiaa Mostafa

Text Advertising is an important field that can affect online advertising. Bid term Suggestion gains a great interest by advertisers. Advertisers should choose carefully the bid terms and the relevant phrases that represent their products or services, otherwise online users would not click on the ad that transfer them to the purchase page of the product or the service being promoted. The paper proposes a method that suggests the keywords of a web page based on the frequent terms in a web page while including the lexical relationship (synonyms) of these words. An experiment is executed to validate the method while the method's result is comparable with Google ad word tool. The proposed method had shown better results than working only with the frequency of the terms.

Entropy Based Sensitivity Analysis and Visualization of Social Networks

Ekrem Serin, Selim Balcisoy

This paper introduces a technique to analyze and visualize social networks using Shannon's entropy model. Entropy is exploited to measure the information amount in social network graphs, and to conduct sensitivity analyses. Novel measures such as degree, betweenness and closeness entropies are evaluated to find the change in graph entropy or the actors. In this work we present a visualization approach that uses coloring, sizing and filtering to help the users perceive the communicated information. The result of sensitivity analyses is integrated into the visualization using the change amount caused by the actors as information. The main contribution of this study is a visualization where the information communicated from a social network is enhanced by the help of sensitivity analyses.

Interlocking Directorates and Profitability: A Social Network Analysis of Fortune 500 Companies

Alton Y.K. Chua, Radhika Shenoy Balkunje

Given the conflicting results from past research, this paper uses social network analysis (SNA) to empirically determine the relationship between interlocking directorates and corporate profitability. Specifically, it seeks to address two research questions: (1) what relationship exists between the number of director interlocks and companies' profitability? (2) what relationship exists between power within the network of interlocking directorates and companies' profitability? The dataset was drawn from 2010 Fortune 500 companies. Regression models were developed to address the two research questions. To mitigate the effects of confounding factors, three control variables, namely, size, age and sector of the companies were used. Results suggest that both interlocks and power asserted a positive linear relationship with companies' profitability. However, the control variables did not seem to influence these relationships. Moreover, prediction of companies' profitability was found to be a complex undertaking. In conclusion, this paper acknowledges a few limitations and offers suggestions for future research.

A Novel Search Engine Based on Social Relationships in Online Social Networking Website

Hsiao-Hsuan Lu, I-Hsien Ting, Shyue-Liang Wang

In recent years, social networking sites have becoming important platforms for users to establish the relationships between each other. As time goes by, the links between people will form the so-called ¡§Strong Links¡¨. For those users, information provided by the friends with strong link is considered as more interesting and useful. Most of recent search engines are designed based on only measuring the similarity between keywords and articles. However, the social relations between authors of articles and searcher have not been taken into

account in recent research. Therefore, in order to improve the performance of recent search engines, we include the measurement of social relationships in search engine and expect the search quality can be improved. In this study, we collected the data from Facebook to calculate the social relationship. About the content, the data will be processed by using CKIP (Chinese word net) and TF-IDF. Finally, we combine key-word frequency and social relations as a value, which is called the Social Ranking vaule. The value will be used as the key to rank the search results. In this paper, we will also demonstrate a real example to explain the proposed methodology as well as a system interface.

Exploring Imagery-driven Service Framework on Social Network Service

Yuan-Chu Hwang, Wei-Cheng Shiau

Service science aims to seek reciprocal value creation to increase the productivity, quality, and opportunities of innovation. How social network services (SNS) could aggregate the energy power to link the needs of customers and shape the designs for better service utilization is an important issue in social network research. In this paper, we concentrate on the imagery-driven innovation approach that highlights how the service imagery could influence the user behaviors and activities in the social network services. An imagery framework for social network services is proposed that explore the opportunities and encourage imagery-driven innovation on SNS. The SNS imagery framework also provides a blueprint to establish service roadmap for future social network services. Qualitative analyses on multiple cases are studied to support the SNS imagery framework.

Web Data Management through Crowdsourcing Upon Social Networks

Marco Brambilla, Alessandro Bozzon

Retrieval and management of Web data is becoming a more and more complex problem, due to the amount of information to be dealt with, to the diversity of the information sources and of the data formats, and to the evolving expectations of users. In particular, some tasks such as quality assessment, opinion making, and sense extraction cannot be completely delegated to automatic procedures. More and more users are increasingly relying on social interaction to complete and validate the results of their online activities. For instance, scouting “interesting” results, or suggesting new, unexpected search directions in information seeking processes occurs in most times aside of the search systems and processes, possibly instrumented and mediated by a social network. In this paper we propose paradigm that embodies crowds and social network communities as first-class sources for the information management and extraction on the Web. Our approach aims at filling the gap between traditional Web systems (CMS, search engines and others), which operate upon world-wide information, with social systems, capable of interacting with real people, in real time, to capture their opinions, suggestions, and emotions by

leveraging crowd sourcing practices and making them viable upon a social network. This enormously enriches the data manipulation experience for the user can be enormously enriched.

An Upper Ontology for the Social Web

Jens Grabarske, Dominic Heutelbeck

Ontologies modeling social graphs often use FOAF as an upper ontology. However, this approach has significant drawbacks. This paper analyzes the requirements for upper social ontologies and illustrates the shortcomings of FOAF and proposed extensions. An alternative approach is presented which is suspected to be sufficient to solve the identified problems.

An Evolutionary-Based Method for Reconstructing Conversation Threads in Email Corpora

Mostafa Dehghani, Masoud Asadpour, Azadeh Shakery

Email is a type of Web data which is produced in enormous quantities. It is beneficial to detect conversation threads contained in the email corpora for various applications, including discussion search, expert finding and even email clustering and classification. Conversation thread in email corpora can be defined as a cluster of exchanged emails among the same group of people by reply or forwarding on the same topic. According to this definition, we can define parent-child relation between emails, so email conversation threads seem to demonstrate tree structure. This paper presents a new approach based on genetic programming for reconstruction of conversation threads in emails data. This approach considers finding email conversation threads as an optimization problem, and exploits genetic programming to search intelligently in the space of possible solutions. Rather than several studies that have been conducted on this problem, this work concentrates on detecting accurate structure of conversation threads in high recall. This paper provides a comprehensive evaluation on the BC3 data set. Preliminary results suggest that our method provides acceptable precision and higher recall than existing methods.

Blackboard Data Spaces for the Elicitation of Community-based Lightweight Ontologies

P. Cristian Vasquez

In open information systems, users can act as consumers and producers of meaningful content. They can have an active role in capturing the semantics that they use within their communities. This approach use the notion of web blackboards, which provide a playground for participants to reach better agreements about a some topic. Multiple blackboards can be connected, forming networks where the content within a blackboards automatically spread. In this approach, the blackboards might converge or diverge in distinct variants, providing a model where

distinct local agreements or ontology views can co-exist within a system. The social construction dynamics benefit from mapping analysis and pattern observation, aiming to increase the underlying community awareness. In the same way, these dynamics are persisted, aiming to capture the community shared understanding across time.

An Efficient Map-Reduce Algorithm for the Incremental Computation of All-Pairs Shortest Paths in Social Networks

Sushant S. Khopkar, Rakesh Nagi, Alexander G. Nikolaev

Today's social networks are getting larger, and the need to analyze datasets with millions of nodes and billions of edges is not uncommon any more. As a network of social relationships evolves by the addition of new nodes and edges, fast algorithms are desirable for the recomputation of key network measures such as actor centrality. The distributed computing paradigm offers a scalable approach to addressing the recomputation challenge. This paper develops a Map-Reduce implementation of an incremental All-Pairs Shortest Path (APSP) algorithm. The incremental nature of the approach allows for performing minimal work in updating centrality measures, while the Map-Reduce implementation makes it scalable to large data. The key idea of the incremental APSP algorithm [1] is based on the efficient use of past information about the shortest paths between any node and the neighbors of the newly added node. A presented parallelized version of the algorithm relies on a three-step iterative execution of the "map" and "reduce" jobs. Experiences with its implementation are reported in application to a real-world dataset containing 7115 nodes. The experimental runs were performed using the Amazon's EMR service.

Density-based Community Identification and Visualisation

Michal Kozielski, Wojciech Filipowski, Dominik Popowicz, Lukasz Warchal

Community can be generally defined as a sub graph where nodes are more densely connected with each other than with the rest of a network. Such definition makes application of density-based clustering methods to community identification justified and natural. Moreover, density-based methods have many extensions enabling their application to complex data analysis. Therefore, the analysis of the characteristics of density-based clustering methods in application to community identification is important and valuable. The article presents and evaluates new similarity measures that can be utilised by the approaches to density-based community identification. Several experiments on real life and generated networks are performed to show and explain the differences between these measures and to compare them with other methods. The results show that the new measures improve the quality of analysis and that density-based clustering algorithms can be valuable community identification methods.

Detecting Social Bookmark Spams Using Multiple User Accounts

Yuta Sakakura, Toshiyuki Amagasa, Hiroyuki Kitagawa

This paper proposes a scheme of detecting gIntensive Bookmarking using Multiple Accountsh (IBMA), where many social bookmark accounts are used to create bookmark entries linking to the target web resources with the aim of increasing site visitors or optimizing search result ranking. To efficiently detect IBMA, we propose to use clustering social bookmark user accounts according to the similarity with respect to the book marked web resources or web sites. Specifically, we cluster users who create bookmarks linking to similar set of web resources or web sites. For this, we propose three similarity measurements over two sets of bookmarks. We experimentally show that the proposed scheme successfully detects IBMA spammers in a real dataset. We also evaluate the accuracy of the proposed scheme with varying the similarity measurements, and characterize them.

Applying Trust Metrics Based on User Interactions to Recommendation in Social Networks

Alberto Lumbreras, Ricard Gavaldà

Recommender systems have been strongly researched within the last decade. With the arising and popularization of digital social networks a new field has been opened for social recommendations. Considering the network topology, users interactions, or estimating trust between users are some of the new strategies that recommender systems can take into account in order to adapt their techniques to these new scenarios. We introduce MarkovTrust, a way to infer trust from Twitter interactions and to compute trust between distant users. MarkovTrust is based on Markov chains, which makes it simple to be implemented and computationally efficient. We study the properties of this trust metric and study its application in a recommender system of tweets.

Method of Visualizing Relations between Tweets to Facilitate Discussions via Twitter

Yasuhiro Yamada, Akira Hattori, Tasuku Kobayashi, Haruo Hayami

It is common for clients using Twitter to adopt displays that indicate their posts (tweets) as textboxes, and they are individually and perpendicularly lined up in order. When several people discuss a particular topic, they frequently post tweets as "replies". However, it is difficult for users in these types of displays to obtain the information that is required and beneficial to understand to make progress in particular discussions. Consequently, we propose a system, which we evaluated, of visualizing the relations between tweets by using digraphs.

A Method for Local Community Detection by Finding Core Nodes

Tiantian Zhang, Bin Wu

Currently, the detection of global community structure in networks has gathered a lot of attention. Most of the methods need global knowledge of the graphs which would be unrealistic to get when the graphs are too large or evolve too quickly. Moreover, sometimes we are only interested in the community structures of some given nodes, not all nodes. So detecting the community of a given node i.e. local community detection is more appropriate. Most of the proposed solutions for local community detection built upon the source nodes are sensitive to the position of source nodes. In this paper, we propose a method to detect local community of a given node by finding the core node of the community firstly. Then expand the core node's cliques to get community of the given node. We validate our method on real-world networks whose community structures are available. The result shows that our method can get high recall and precision score and is quite effective and flexible to identify local communities, irrespective of the source node position.

Early and Late Fusion Methods for the Automatic Creation of Twitter Lists

Mengjiao Wang, Donn Morrison, Conor Hayes

Twitter's list feature allows users to organize their followees into groups for easier information access and filtering. However, the percentage of users using lists is very small and most existing lists have only a few members. One reason for this may be that curating groups of Twitter users is a time consuming task. In this paper, we propose early and late fusion methods for automatically clustering followees using both graph structure and tweet content. We evaluate our approaches using ground-truth Twitter lists crawled via the Twitter API and show that the late fusion method outperforms both the baselines and the early fusion method.

Influence of Feedback from SNS Members on Consumer Behavior in Electronic Commerce

Vanessa R. Bracamonte Lesma, Hitoshi Okada

Positive feedback from previous users can affect consumer behavior towards an electronic commerce website. Social Network Sites provide a way to gather feedback from SNS users and show them on websites, but not much research has been done on this type of feedback mechanism. In this study we investigate the influence that SNS-based feedback has on the behavior of consumers, comparing the effect in consumers who are members of a SNS vs. those who are not members. We conducted a survey in Japan, using a mock Thai website that showed three different levels of SNS information. We found that consumers who are SNS members have a more positive attitude towards the website when SNS-based

feedback information is shown. We also found that the nationality of the SNS members giving the feedback affects trust and risk perceptions differently.

Twitter as a Tool for Predicting Elections Results

Juan M. Soler, Fernando Cuartero, Manuel Roblizo

The use of social networking sites, such as Twitter, for various purposes, continues to grown since its first appearance. This social net is a microblogging site to share short messages on a variety of topics. In particular, political elections are a very interesting field to exchange views using this platform. In this paper we have used Spanish elections to investigate the use of Twitter for this purpose, and to find out if the conversations maintained there can anticipate, in some way, the results of the elections. In order to do this, we have developed a tool, called Tara tweet, to define experiments and to capture the defined conversations, and have applied it to the cases of three Spanish elections during 2011 and 2012. Our results show that Twitter is used for political discussion, and that the references to the different political parties correlate, significatively, with the votes of the electors. This is an indicator that Twitter may be used by social researchers as a tool, among others, to predict future results of the elections. Of course, with due caution because the measured data correspond to distinct actions, so obviously, much more research and studies should be done in this field.

Identification of Group Changes in Blogosphere

Bogdan Gliwa, Stanislaw Saganowski, Anna Zygmunt, Piotr Brᴓdka, Przemyslaw Kazienko, Jaroslaw Kozak

The paper addresses a problem of change identification in social group evolution. A new SGCI method for discovering of stable groups was proposed and compared with existing GED method. The experimental studies on a Polish blogosphere service revealed that both methods are able to identify similar evolution events even though both use different concepts. Some differences were demonstrated as well.

Evolution of Author's Topic in Authorship Network

Sarka Zehnalova, Zdenek Horak, Milos Kudelka, Vaclav Snasel

There may be several reasons why people publish together. Above all, the fact that the authors share common professional interests is the main reason. In our research we work with the DBLP dataset which contains the basic bibliographic information of publications from the computer science field. These data are freely available and contain highly relevant information about publication activity from the period of nearly fifty years, even though they are not complete. One of the goals of our research is to analyze and visualize the evolution of authors and co-

authorship from the point of view of research topics. We present the results of our research in this paper. One of the results is also visualization in our online FORCOA.NET system.

CoKIM: Collaborative and Social Knowledge-Based Incident Manager

José Luis López-Cuadrado, Israel González-Carrasco, Ricardo Colomo-Palacios, Rodrigo Gómez-Rodríguez, Ángel García-Crespo

Dealing with user incidents, questions and problems are important tasks during the maintenance of the software. It is specially important for service providers, which must provide the users with a quick and reliable answer. Many times the incident has been previously solved for other users. In these cases, an appropriate knowledge management strategy could help to improve response times as well as the accuracy of such responses. This paper presents the work in progress of Co KIM, a knowledge based framework for incident management, based on the collaboration between users and experts. The framework is able to propose solutions for incidents based on the previous responses for other similar problems. The selection of the solutions is based on the previous responses from experts and the rating of experts, clients and responses.

Using Location-Based Social Networks to Validate Human Mobility and Relationships Models

T. Nguyen, B.K. Szymanski

We propose to use social networking data to validate mobility models for pervasive mobile ad-hoc networks (MANETs) and delay tolerant networks (DTNs). The Random Waypoint (RWP) and Erdos-Renyi (ER) models have been a popular choice among researchers for generating mobility traces of nodes and relationships between them. Not only RWP and ER are useful in evaluating networking protocols in a simulation environment, but they are also used for theoretical analysis of such dynamic networks. However, it has been observed that neither relationships among people nor their movements are random. Instead, human movements frequently contain repeated patterns and friendship is bounded by distance. We used social networking site Go Walla to collect, create and validate models of human mobility and relationships for analysis and evaluations of applications in opportunistic networks such as sensor networks and transportation models in civil engineering. In doing so, we hope to provide more human-like movements and social relationship models to researchers to study problems in complex and mobile networks.

Four Degrees of Separation, Really

Paolo Boldi, Sebastiano Vigna

We recently measured the average distance of users in the Facebook graph, spurring comments in the scientific community as well as in the general press (Four Degrees of Separation). A number of interesting criticisms have been made about the meaningfulness, methods and

consequences of the experiment we performed. In this paper we want to discuss some methodological aspects that we deem important to underline in the form of answers to the questions we have read in newspapers, magazines, blogs, or heard from colleagues. We indulge in some reflections on the actual meaning of "average distance" and make a number of side observations showing that, yes, 3.74 "degrees of separation" are really few.

Enterprise Wisdom Captured Socially

Charalampos Chelmis, Vikram Sorathia, Viktor K. Prasanna

Data availability in online social networks as well as the business world has lately not been an issue. Vast amounts of data are being generated by social networking users in the form of informal interactions. What has been an issue, is the transformation of data into useful information, that in time and with appropriate processing becomes knowledge. In this paper we examine knowledge generation under informal social communications, based on semantically enriched user-generated data and associated metadata. We dynamically capture users' interests and expertise using such semantically enriched content. Knowledge networks of users emerge, exhibiting collective intelligence. To capture such collective knowledge, we propose a novel knowledge base paradigm, which seamlessly integrates information from multiple platforms and facilitates knowledge extraction, mining, discovery and inferencing. Using semantically enriched user profiles, we compute semantic similarity between users and content in a joint semantic space, driving numerous applications.

Real Time Distributed Community Structure Detection in Dynamic Networks

Valerie Galluzzi

Communities can be observed in many real-world graphs. In general, a community can be thought of as a portion of a graph in which intra-community links are dense while inter-community links are sparse. Automatic community structure detection has been well studied in static graphs. However, many practical applications of community structure involve networks in which communities change dynamically over time. Several methods of detecting the community structure of dynamic graphs have been proposed, however most treat the dynamic graph as a series of static snapshots, which creates unrealistic assumptions. Others require large amounts of computational resources or require knowledge of the dynamic graph from start to finish, relegating them to post-processing. For those who desire real-time community structure detection distributed over the observing network, these solutions are insufficient. This paper proposes a new method of community structure detection which allows for real time distributed detection of community structure.

Constructing Professional Resource Networks from Career Biographical Database

Renuka Hodigere, Diana Bilimoria

Advances in technology and its usage have resulted in vast quantities of information being available on the public domains accessible via the internet. In this study, we explore the viability of one method of using information that has already been collated, to construct a network of the professional resources of a population not easily accessible viz., corporate directors.

Combining Relations and Text in Scientific Network Clustering

David Combe, Christine Largeron, Elⵁd Egyed-Zsigmond, Mathias Géry

In this paper, we present different combined clustering methods and we evaluate their performances and their results on a dataset with ground truth. This dataset, built from several sources, contains a scientific social network in which textual data is associated to each vertex and the classes are known. Indeed, while the clustering task is widely studied both in graph clustering and in non supervised learning, combined clustering which exploits simultaneously the relationships between the vertices and attributes describing them, is quite new. We argue that, depending on the kind of data we have and the type of results we want, the choice of the clustering method is important and we present some concrete examples for underlining this.

Investigation of Policy Structure in Legal Authorities of Institutions Related to Water Policy Making and Management in Iran through Network Analysis

Hannaneh Mohammadi Kangarani, Sima Rafsanjani Nezhad, Sara Rafsanjani Nezhad, Javad Bakhshi

The Codified development documents of Iran indicate that the planned objectives in water section have not been successfully operated. Therefore, the present research has tried to find the basic reasons of such failure. The study explored the power structure in legal network of formal institutions connected with water management in the framework of "New Institutionalism Theory". Network of legal authorities of formal institutions related to water resources in national level was delineated with "Network Analysis" through UCINET and Net draw Software's. The data have been analyzed using the different indices in network analysis such as centrality, degree, power, and cut point. The results show that there is no good power distribution among water section institutions and no firm relationship among administrative and authoritative institutions in this section. Moreover, nongovernmental institutions do not profit from an appropriate position

in these documents. So, it is necessary for policy makers to pay more attention to a balanced power distribution among institutions in national and provincial levels.

Studying Diffusion of Viral Content at Dyadic Level

Anita Zbieg, Blazej Zak, Jaroslaw Jankowski, Radoslaw Michalski, Sylwia Ciuberek

Diffusion of information and viral content, social contagion and influence are still topics of broad evaluation. As theory explaining the role of influentials moves slightly to reduce their importance in the propagation of viral content, authors of the following paper have studied the information epidemic in a social networking platform in order to confirm recent theoretical findings in this area. While most of related experiments focus on the level of individuals, the elementary entities of the following analysis are dyads. The authors study behavioral motifs that are possible to observe at the dyadic level. The study shows significant differences between dyads that are more vs less engaged in the diffusion process. Dyads that fuel the diffusion proccess are characterized by stronger relationships (higher activity, more common friends), more active and networked receiving party (higher centrality measures), and higher authority centrality of person sending a viral message.

Group Recommender Model for Boosting and Optimizing Customer Purchases

M. Saravanan, G. Prasad, M. Jagadeesan, Revathy Raman, S. Rekha

Group recommender systems generate a set of recommendations that will satisfy a group of customers with potentially competing purchase interests. This paper proposes a research and operational model which effectively enhances Group Recommender Model to boost the customer purchases. For this purpose, it uses the communication and collaboration of two major sources namely Mobile Money Operator and Outlet. MMO proactively monitors the spending pattern of the customers who make purchases using their mobile money. Outlet performs customer segmentation based on RFM (Recency, Frequency and Monetary) score after which a Recursive Cluster Elimination is performed that eliminates customers within the targeted segment. Recursive Frequent Item set Mining and Recursive Market Basket Analysis are performed for the rest of customers in the targeted segment. From the obtained results, the product preferences of the remaining customers in the segment are identified based on which offers are formulated and recommended for the entire segment. It is then communicated to the MMO that intimates these offers to the potential customers among the segment. This model results in boosting customer purchases, expanding customer base and effects in the profitability of the combined source.

An Interests Discovery Approach in Social Networks Based on Semantically Enriched Graphs

Akram Al-Kouz, Sahin Albayrak

Studying the text messages of a user such as his posts in Facebook or his tweets in Twitter can help in detecting his topics of interests. User in Social Network Systems (SNS) posts text messages about a wide diverse of topics. Posts usually written in a non-standard language, which make it not applicable to the standard Natural Language Processing (NLP) techniques used to catch the relations between words in text. In many cases there are semantic relations between the contained entities of posts that can infer the interest of the user. Bag-Of-Words (BOW) based text classification techniques classify this kind of messages to a wide diverse of topics, but they fail in catching the implicit semantic relation between the contained entities. In this paper we propose a technique to discover the implicit semantic relations between entities in text messages, which can infer the interests of a user. The proposed technique based on a semantically enriched graph representation of entities contained in text messages generated by a user, a new algorithm (Root-Path-Degree) is invented and used to find the most representative sub-graph that reflects the semantic implicit interests of the user. An evaluation was done using manually annotated posts of 687 Facebook users. Precision and Recall results showed our technique performs better than the standard BOW technique.

Leveraging Social Networks to Improve Service Selection in Workflow Composition

John McDowall, Larry Kerschberg

In recent years, social media has expanded from a niche application with a student-focused user base to a mainstream tool used by individuals and business to maintain and expand their social and business networks. The result is a rich source of data that can help service consumers and service providers connect with each other in new and interesting ways. This paper describes research that leverages social media to help users optimize service selections when composing executable workflows from among both web services and physical services.

Identifying Influential Taggers in Trust-Aware Recommender Systems

Sanjog Ray

Trust-ware recommender systems provide the features of personalized product and service recommendations in web based social networks by using the trust connections existing between users and preferences data available for each user. One of the main sources of user preferences data are the tags that users apply to different items. Encouraging users to apply more tags is one of the challenges faced by most social network sites. In this paper we purpose an approach to identify influential taggers in a trust based social network so that efforts to encourage tagging can

be achieved by designing incentives for motivating the influential taggers to apply more tags. In our proposed approach, for every user his tagging influencer is that user in his personal network who has influenced his tagging behavior the most. We define an active user tagging actions has been influenced by a user in his personal network only when the active user tags an item after his influencer has tagged it. The influential taggers in the overall social network are those who have the influenced the maximum number of users in the network. We analyze the real life dataset of Last.fm to show that our approach is different from the current approach of defining those users who have tagged the maximum number of items as the influential users. We also discuss the implications of using our approach.

Clustering Social Networks to Remove Neutral Nodes

Fatemeh Hendijani Fard, Behrouz H. Far

Multi agent systems with autonomous interaction, negotiation and learning capabilities can efficiently model social behavior of individuals participating in a social network. A central problem in a social network is to identify the nodes that actively participate in the expansion of the net both physically and functionally. Several metrics have already been proposed to identify those hot spots. The algorithms to identify hot spots are either heuristic based or computationally expensive. In this paper we use an agent model of the social net and propose a method that can identify the neutral nodes, i.e. the nodes that can never be considered as hot spot nodes given the network topology and rules of negotiation among nodes. Therefore these nodes can be eliminated from the net. A direct advantage of this method is reducing the computational complexity for the configuration and identification of hot spots. Through a case study we have shown that the proposed method can lead to 33% reduction of computation regarding the number of agent types in the example.

Obstacles Incentivize Human Learning: A Network Theoretic Study

Amitash Ramesh, Soumya Ramesh, Sudarshan Iyengar, Vinod Sekhar, C. Pandu Rangan

The current paper is an investigation towards understanding the navigational performance of humans on a network when the`landmark' nodes are blocked. We observe that humans learn to cope up, despite the continued introduction of blockages in the network. The experiment proposed involves the task of navigating on a word network based on a puzzle called the word morph. We introduce blockages in the network and report an incremental improvement in performance with respect to time. We explain this phenomenon by analyzing the evolution of the knowledge in the human participants of the underlying network as more and more landmarks are removed. We hypothesize that humans learn the bare essentials to navigate unless we introduce blockages in the network which would whence enforce upon them the need to explore newer ways of navigating. We draw a parallel to human problem solving and postulate that obstacles are catalysts for humans to innovate techniques to solve a restricted variant of a familiar problem.

Pareto Optimal Allocation in Multi-agent Coalitional Games with Non-linear Payoffs

Usha Sridhar, Sridhar Mandyam

A fully connected network of multiple interacting agents modeled as a cooperative game to attain a common objective has found wide applications in the real world. Competitors frequently come together to work in coalitions that are mutually beneficial to them all, though the allocation of the mutual gains achieved is seldom easy. Shapley value is a popular way to compute payoffs in cooperative games where the agents are assumed to have deterministic, risk-neutral (linear) utilities. This paper explores a class of Multi-agent constant-sum cooperative games where the payoffs are random variables. We introduce a new model based on Borch's Theorem from the actuarial world of re-insurance, to obtain a Pareto optimal allocation for agents with risk-averse exponential utilities. This allocation problem seeks to maximize a linear sum of the expected utilities of a set of agents and the solution obtained at this optimal value naturally maximizes the social welfare of the grand coalition. The four main axioms of the Shapely Value, namely, nullity, additivity, symmetry and efficiency are satisfied by this solution. We show the correspondence of our solution to the Shapley value. As a result we can directly obtain the Shapley value from the allocation values obtained at the Pareto optimum as the individual utility achievements of the grand coalition.

A Navigation Algorithm Inspired by Human Navigation

Vijesh M., Sudarshan Iyengar, Vijay Mahantesh SM, Amitash Ramesh, C. Pandurangan, Veni Madhavan

Evolution of User Activity and Community Formation in an Online Social Network

Andreas Kalaitzakis, Harris Papadakis, Paraskevi Fragopoulou

The paper performs an empirical study of the My Space Online Social Network (OSN). It aims to capture the evolution of user population, to examine user activity, and finally to characterize community formation using two well established community finding algorithms, namely the Fortuna to et al. and the Clique Percolation algorithms. Both algorithms are known to be effective in identifying communities in large graphs, starting at seed nodes and utilizing only local interactions between nodes. One million user profiles were randomly collected in a month's period. For each profile certain attributes were fetched: profile status (public, private, invalid), member since and last login dates, number of friends, number of views, etc. The profiles and their attributes were analyzed in order to reveal the evolution in user population and the activity of the participating members. Significant conclusions were drawn for the synthesis of the population based on profile status, the number of friends, and the duration My Space members stay active. Subsequently, a large number of communities were identified aiming to reveal the structure of the underlying social network graph. The collected data were further analyzed in order to characterize community size and density but also to

retrieve correlations in the activity among members of the same community. A total of 171 communities were detected with Fortunato's algorithm, while using Clique Percolation this number was 201. Results demonstrate that My Space members tend to form dense communities. For the first time, strong correlation in the last login date (the main attribute that shows user activity) for members of the same community was documented. It was also shown that members participating in the same community have similar values for other attributes like for example number of friends. Lastly, there is strong evidence that participation of users in communities inhibits them from abandoning My Space.

Cyber Politics: Understanding the Use of Social Media for Dissident Movements in an Integrated State Stability Framework

Daniel Goldsmith, Michael Siegel

Recent events in North Africa and the Gulf States have highlighted both the fragility of states worldwide and the ability of coordinated dissidents to challenge or topple regimes. The common processes of 'loads' generated by dissident activities and the core features of state resilience and its 'capacity' to withstand these 'loads' have been explored in the traditional "real world" view. More recently, however, there has been increased attention to the "cyber world"—the role of cyber technologies in coordinating and amplifying dissident messages, as well as in aiding regimes in suppressing anti-regime dissidents. As of yet, these two views (real and cyber) have not been integrated into a common framework that seeks to explain overall changes in regime stability over time. Further, emerging uses of social media technologies, such as Twitter have not fully been examined within an overall framework of state stability that represents the nature and dynamics of 'loads' generated by dissident activities in the real (i.e. protests) and cyber (i.e. planning and coordination via cyber venues) domains.

Community Learning from External Information Sources

Sridhar Mandyam, Usha Sridhar

We model the persuasive effect of external information sources such as media on social networks using a new endogenous social learning framework. The agents are thought to hold uninformative probabilistic prior beliefs about an issue that concerns them and learn about this state of the world through a non-Bayesian myopic DeGroot-style update process applied on the priors using social influence 'mixtures'. We model external information sources in this framework as entities that can bring to the attention of agents 'global' beliefs that are potentially from beyond the confines of a community, and may well be in conflict among themselves. In our model agents score these information sources on the basis of how closely the beliefs propounded by the sources match their own beliefs, but determine how to assimilate such beliefs on the basis of the views of their community of connected neighbors. This form of social learning of external information allows local social influences to carry shared views resulting in the potential emergence of modified homophyllic structures, for example to capture the

notion that those who view external information sources in a similar manner might be inclined to demonstrate higher affinities among themselves. We show that this form of social learning of externally expounded beliefs has a learnable dynamic which achieves convergence, and can mirror scenarios where external sources can bring about consensus among opposed cliques, or break emerging consensus. We illustrate the working of the learning model on a simple example.

Landmark Identification in Complex Networks

Suhas Venkatesh, Amitash Ramesh, Udaya Shyama, Sudarshan Iyengar

Identification of landmark nodes in complex networks has been of great interest but the parameters based on which these nodes are selected have been different in different cases. We believe that Landmarks are those nodes that occur most frequently while we explore or navigate a network. We propose 2 methods and efficient algorithms to identify these nodes. We identify these nodes using centrality measures and random walks. We show that the methods proposed are both, effective and efficient. Our approach has been extensively evaluated using synthetic environments. The results obtained verify its robustness along with its applicability.

Web Page Prediction by Clustering and Integrated Distance Measure

G. Poornalatha, Prakash S. Raghavendra

The tremendous progress of the internet and the World Wide Web in the recent era has emphasized the requirement for reducing the latency at the client or the user end. In general, caching and prefetching techniques are used to reduce the delay experienced by the user while waiting to get the web page from the remote web server. The present paper attempts to solve the problem of predicting the next page to be accessed by the user based on the mining of web server logs that maintains the information of users who access the web site. The prediction of next page to be visited by the user may be pre fetched by the browser which in turn reduces the latency for user. Thus analyzing user's past behavior to predict the future web pages to be navigated by the user is of great importance. The proposed model yields good prediction accuracy compared to the existing methods like Markov model, association rule, ANN etc.

Fraud Detection: Methods of Analysis for Hypergraph Data

Anna Leontjeva, Konstantin Tretyakov, Jaak Vilo, Taavi Tamkivi

Hyper graph is a data structure that captures many-to-many relations. It comes up in various contexts, one of those being the task of detecting fraudulent users of an on-line system given known associations between the users and types of activities they take part in. In this work we explore three approaches for applying general-purpose machine learning methods to such data. We evaluate the proposed approaches on a real-life dataset of customers and achieve promising results.

On the T-graph of a Commutative Ring

Ahmad Abbasi

Let R be a commutative ring with non-zero identity and let I be a proper ideal of R. In this talk we study the T-graph consisting of all elements of R as vertices, such that two vertices x and y in the graph are adjacent if and only if, where S(I) = {a epsilon R | ra epsilon, for somer epsilon R - I}.

Recommendation in Online Health Communities

Steven P. Crain, Ke Zhou, Hongyuan Zha

Online health communities enable patients to share ideas and experiences to improve their understanding of themselves and their treatment. Personalized recommendation helps patients find appropriate online social resources and protects and enhances the character of the resources. Recommendation needs to leverage a wide variety of types of interactions, but most state-of-the-art recommender systems using matrix factorization are designed for one type of relation with a restricted range of ratings. In this paper, we address the problem of recommendation using multiple relations with unrestricted ranges of ratings derived from social behaviors. We model multiple relations through multi-task matrix factorization where the latent profiles are \emph{partially shared} between relations. Moreover, we offer two new techniques for transforming between the distribution produced by latent factor models and power-law-distributed ratings. The experiments conducted over a data set from a diabetes community suggest that the proposed model can improve the accuracy of recommendation in online health communities. The performance for recommendation to users is very good (mean rank of test examples is less than 5 out of 30) but poorer for recommendation to groups (mean rank 10 out of 30).