State of the art
Description of the dynamics

DynGraph project
ANR-10-JCJC-0202

March 9, 2011

1 General description

Most of the works dealing with the description of dynamic networks have been concerned with the in-depth study of specific real-world dynamic graphs. The case of contact or proximity networks between individuals, measured via electronic sensors or bluetooth cellphones for instance, has received much attention [8, 9, 11, 46, 51]. Other cases include P2P exchange networks [19, 31, 32], the internet's topology [29, 38, 36, 41, 40, 27], biological networks [1, 2, 42, 53, 47], citation networks [33], and different types of online social networks [13, 48].

Some works have focused on general questions, relevant for the study of any dynamic graph. These questions include algorithmic methods for manipulating dynamic graphs [7, 23, 24, 15], the study of the evolution of formation of specific patterns [54, 10, 28, 43, 25, 14] or the evolution of distances in the network [34], or dynamic graph drawing [17, 18].

Some works study the time behavior of nodes in a dynamic networks, i.e. the times at which nodes create new links, to detect generic behavioral patterns, and/or nodes with anomalous behavior, and/or times at which the global dynamics of the network changes [26, 33].

Some surveys on the topic have been written [3, 37].

2 Link prediction

Link prediction is a key research problem in network dynamic analysis. Several works study this problem. Most of them are based on measures of similarity between nodes. For instance, in [35] the authors examine several topological measures (such as Jaccard coefficient, Adamic/Adar coefficient, SimRank, etc.) based on node neighborhoods and the set of all paths between nodes. They use these measures for ranking possible future co-authors collaborations. In [21] the author proposes to use another topological measure called generalized clustering coefficient. In [20, 39] the authors add several non-topological measures based on node attributes (such as keyword match, number of papers, geographic proximity, KL-divergence of two nodes' topic distribution, etc.) and they use a supervised learning algorithm to perform link prediction. In a similar way, the authors of [5] predict co-authoring of publications by using topological measures computed in the co-authoring graph and indirect topological measures computed using the co-author graph (where two papers are linked if they are signed by a same author). The authors of [55] add another measure (local probabilistic model) to estimate the co-occurrence probability of two nodes, and in [52] the authors extend this by incorporating available temporal information. Finally, the authors of [12] use hierarchical decomposition of the social network and use it for predicting missing links. Authors generate a set of hierarchical random graphs, and they compute average probability of connection between two nodes within these hierarchical random graphs.

Some works address the topic of link prediction in bipartite networks. In [22], the authors adapt some topological measures used in classical graphs. For the measures based on node neighborhoods between two nodes (u, v), the authors consider neighbors of u in the bipartite graph and set of neighbors of v. For the measures based on the set of all paths between nodes, they compute directly in the bipartite graph. To go further, the authors of [4] consider two transformations of the bipartite graph into a classical one, and they use two indirect measures. For predicting link (u, v), they consider the classical graph containing u, and they apply the topological measures between u and the neighbors of v in the bipartite graph, and conversely.

3 Bias in the observation of the dynamics

Finally, a number or works have addressed the question of the reliability of the observed properties of a dynamic network. They studied possible measurement biases occurring in this context.

One bias comes from the fact that, because we can only observe a system for a finite period of time, events occurring before or after the observation window are missed. This has mainly been acknowledged for *churn*, i.e. the dynamicity of users, in P2P systems [6, 44, 45, 56, 50, 49], but also in other contexts [25, 30].

Willinger et. al. [57] addressed, in the context of IP flows, the question of whether the observation window is long enough to characterize some dynamic properties. They study the standard deviation of the flow size distribution as a function of the measurement length, and argue that the fact that it does not converge means that the samples may come from an underlying distribution with infinite variance. This in turn may make it difficult to fit the observed properties with a model.

The *create-based method* [44, 45] is based on the observation that being able to only capture accurately the length of sessions that begin and end within the measurement window creates a bias towards short sessions. To remove this bias, the measurement window of length T is divided into two halves, and only the sessions that begin during the first half and last less than $T/2$ are considered. This leads to an unbiased estimation of sessions with length less than $T/2$. This method only applies to properties for which a notion of session can be defined, which is not always the case.

Finally, the bias caused by the finiteness of the observation window is not the only one occurring in dynamic networks. Stutzbach and Rejaie [50] studied different aspects of peer dynamics in three different classes of P2P systems

(Gnutella, Kad and BitTorrent). They carefully analyzed the different kinds of bias that may influence such a study, and presented a list of those they identified, which includes problems linked to accurate peer identification.

Wang et. al. [56] argue that the create-based method is biased when the data is obtained through periodic sampling, because short events may be missed or incorrectly observed. They propose a new sampling algorithm called RIDE (ResIDual-based Estimator) which measures session length distributions with high accuracy and requires a low sampling frequency.

Stutzbach et. al. [49] investigate the issues arising when the whole system is not known, and informations about the nodes and links are obtained by a sampling procedure (in this case, random walk-based methods), in the case where the system evolves while the sampling process is under progress.

Friggeri et. al. [16] studied contact networks captured with sensors able to detect when they are close to each other. They studied the bias on the observed contact duration caused by the fact that some sensors may fail to detect each other at some times.

References

[1] M.M. Babu, L. Aravind, and S.A. Teichmann. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol.*, 358(2):614–633, 2006.

[2] M.M. Babu, N.M. Luscombe, L. Aravind, M. Gerstein, and S.A. Teichmann. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol.*, 14(3):283–291, 2004.

[3] A.-L. Barabási. The physics of the web. *Physics World*, 14:33–38, 2001.

[4] Nesserine Benchettara, Rushed Kanawati, and Celine Rouveirol. Supervised machine learning applied to link prediction in bipartite social networks. *Social Network Analysis and Mining, International Conference on Advances in*, 2010.

[5] Nesserine Benchettara, Rushed Kanawati, and Céline Rouveirol. A supervised machine learning link prediction approach for academic collaboration recommendation. In *RecSys '10: Proceedings of the fourth ACM conference on Recommender systems*, 2010.

[6] Ranjita Bhagwan, Stefan Savage, and Geoffrey M. Voelker. Understanding availability. In *IPTPS*, pages 256–267, 2003.

[7] B. Bui-Xuan, A. Ferreira, and A. Jarry. Evolving graphs and least cost journeys in dynamic networks. In *Proceedings of WiOpt'03 – Modeling and Optimization in Mobile, Ad-Hoc and Wireless Networks*, pages 141–150, Sophia Antipolis, March 2003. INRIA Press.

[8] R. Calegari, M. Musolesi, F. Raimondi, and C. Mascolo. CTG: A connectivity trace generator for testing the performance of opportunistic mobile systems. In *European Software Engineering Conference and the International ACM SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE07)*, Dubrovnik, Croatia, 2007.

[9] A. Chaintreau, J. Crowcroft, C. Diot, R. Gass, P. Hui, and J. Scott. Pocket switched networks and the consequences of human mobility in conference environments. In *WDTN*, pages 244–251, 2005.

[10] Jeffrey Chan, James Bailey, and Christopher Leckie. Discovering correlated spatio-temporal changes in evolving graphs. *Knowledge and Information Systems*, 16(1):53–96, 2008.

[11] A. Clauset and N. Eagle. Persistence and periodicity in a dynamic proximity network. In *DIMACS Workshop*, 2007.

[12] Aaron Clauset, Cristopher Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 2008.

[13] Jean-Philippe Cointet and Camille Roth. Socio-semantic Dynamics in a Blog Network. In *IEEE SocialCom 09 Intl Conf Social Computing*, pages 114–121. IEEE CS, 2009.

[14] G. Cormode, F. Korn, S. Muthukrishnan, and Y. Wu. On signatures for communication graphs. In *International Conference on Data Engineering (ICDE)*, 2008.

[15] David Eppstein. All maximal independent sets and dynamic dominance for sparse graphs. In *Proc. 16th Symp. Discrete Algorithms*, pages 451–459. ACM and SIAM, January 2005.

[16] Adrien Friggeri, Guillaume Chelius, Eric Fleury, Antoine Fraboulet, France Mentr, and Jean-Christophe Lucet. Reconstructing social interactions using an unreliable wireless sensor network. *Computer Communications*, 34 (5), 2011.

[17] Yaniv Frishman and Ayellet Tal. Online dynamic graph drawing. *IEEE Transactions on Visualization and Computer Graphics*, 14:727–740, 2008.

[18] Carsten Görg, Peter Birke, Mathias Pohl, and Stephan Diehl. Dynamic Graph Drawing of Sequences of Orthogonal and Hierarchical Graphs, 2005.

[19] J.-L. Guillaume, S. Le-Blond, and M. Latapy. Clustering in p2p exchanges and consequences on performances. In *IPTPS*, 2005.

[20] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki. Link prediction using supervised learning. In *In Proceedings of SDM 06 workshop on Link Analysis, Counterterrorism and Security*, 2006.

[21] Zan Huang. Link Prediction Based on Graph Topology: The Predictive Value of Generalized Clustering Coefficient. In *Workshop on Link Analysis Dynamics and Static of Large Networks (LinkKDD'06), in conjunction with KDD-2006*, 2006.

[22] Zan Huang, Xin Li, and Hsinchun Chen. Link prediction approach to collaborative filtering. In *In Proceedings of the Joint Conference on Digital Libraries (JCDL05). ACM*, 2005.

[23] D. Kempe, J. Kleinberg, and A. Kumar. Connectivity and inference problems for temporal networks. In *Proc. 32nd ACM Symposium on Theory of Computing*, 2000.

[24] Valerie King and Garry Sagert. A Fully Dynamic Algorithm for Maintaining the Transitive Closure. *Journal of Computer and System Sciences*, 65(1):150–167, 2002.

[25] Gueorgi Kossinets and Duncan J Watts. Empirical analysis of an evolving social network. *Science*, 311(5757):88–90, January 2006.

[26] Ravi Kumar, Jasmine Novak, Prabhakar Raghavan, and Andrew Tomkins. On the Bursty Evolution of Blogspace. *World Wide Web*, 8(2):159–178, June 2005.

[27] M. Lad, D. Massey, and L. Zhang. Visualizing Internet Routing Changes. *IEEE Transactions on Visualization and Computer Graphics, special issue on Visual Analytics*, 2006.

[28] Mayank Lahiri and Tanya Y. Berger-Wolf. Mining Periodic Behavior in Dynamic Social Networks. In *IEEE International Conference on Data Mining*, pages 373–382. IEEE, 2008.

[29] M. Latapy, C. Magnien, and F. Ouédraogo. A radar for the internet. In *Proceedings of ADN'08: 1st International Workshop on Analysis of Dynamic Networks, in conjonction with IEEE ICDM 2008*, 2008.

[30] Matthieu Latapy and Clémence Magnien. Complex Network Measurements: Estimating the Relevance of Observed Properties. In *2008 IEEE INFOCOM - The 27th Conference on Computer Communications*, pages 1660–1668. IEEE, 2008.

[31] S. Le-Blond, M. Latapy, and J.-L. Guillaume. Statistical analysis of a P2P query graph based on degrees and their time evolution. In *IWDC*, 2004.

[32] Stevens Le Blond, Fabrice Le Fessant, and Erwan Le Merrer. Finding Good Partners in Availability-Aware P2P Networks. In *Stabilization, Safety, and Security of Distributed Systems*, volume 5873 of *Lecture Notes in Computer Science*, pages 472–484, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[33] E. A. Leicht, G. Clarkson, K. Shedden, and M. E. J. Newman. Large-scale structure of time evolving citation networks. *Eur. Phys J. B*, 59, 2007.

[34] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (ACM TKDD)*, 1(1), 2007. Arxiv physics/0603229.

[35] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *Proc. 12th International Conference on Information and Knowledge Management (CIKM)*, 2003.

[36] Clémence Magnien, Frédéric Ouédraogo, Guillaume Valadon, and Matthieu Latapy. Fast Dynamics in Internet Topology: Observations and First Explanations. In *2009 Fourth International Conference on Internet Monitoring and Protection*, pages 137–142. IEEE, 2009.

[37] Mark Newman, Albert-Laszlo Barabási, and Duncan J. Watts. *The Structure and Dynamics of Networks.* Princeton University Press, Princeton, USA, 2006.

[38] Ricardo Oliveira, Beichuan Zhang, and Lixia Zhang. Observing the Evolution of Internet AS Topology. In *Proceedings of ACM SIGCOMM*, 2007.

[39] Joshua O'Madadhain, Jon Hutchins, and Padhraic Smyth. Prediction and ranking algorithms for event-based network data. *ACM SIGKDD Explorations Newsletter*, 2005.

[40] Jean-Jacques Pansiot. Local and Dynamic Analysis of Internet Multicast Router Topology. *Annales des télécommunications*, 62:408–425, 2007.

[41] S.-T. Park, D. M. Pennock, and C. L. Giles. Comparing static and dynamic measurements and models of the Internet's AS topology. In *IEEE Infocom*. IEEE, 2004.

[42] Robert J J. Prill, Pablo A A. Iglesias, and Andre Levchenko. Dynamic properties of network motifs contribute to biological network organization. *PLoS Biol*, 3(11):1881–1892, 2005.

[43] D. Romero and J. Kleinberg. The Directed Closure Process in Hybrid Social-Information Networks, with an Analysis of Link Formation on Twitter. In *Proc. 4th International AAAI Conference on Weblogs and Social Media*, 2010.

[44] D. Roselli, J. R. Lorch, and T. E. Anderson. A comparison of file system workloads. In *Proc. of USENIX Annual Technical Conference*, 2000.

[45] Stefan Saroiu, Krishna P. Gummadi, and Steven D. Gribble. Measuring and analyzing the characteristics of napster and gnutella hosts. *Multimedia Systems*, 9:170–184, 2003.

[46] A. Scherrer, P. Borgnat, E. Fleury, J.-L. Guillaume, and C. Robardet. Description and simulation of dynamic mobility networks. *Computer Network*, 52:2842–2858, 2008.

[47] Ralf Steuer, Adriano Nunes Nesi, Alisdair R. Fernie, Thilo Gross, Bernd Blasius, and Joachim Selbig. From structure to dynamics of metabolic pathways: application to the plant mitochondrial TCA cycle. *Bioinformatics*, 23(11):1378–85, June 2007.

[48] Alina Stoica and Christophe Prieur. Structure of Neighborhoods in a Large Social Network. In *IEEE International Conference on Social Computing (SocialCom-09)*. IEEE, 2009.

[49] D. Stutzbach, R. Rejaie, N.G. Duffield, S. Sen, and W. Willinger. On unbiased sampling for unstructured peer-to-peer networks. *IEEE/ACM Transactions on Networking*, 17(2), 2009.

[50] Daniel Stutzbach and Reza Rejaie. Understanding churn in peer-to-peer networks. In *Internet Measurement Conference*, pages 189–202, 2006.

[51] Pierre Ugo Tournoux, Jérémie Leguay, Marcelo Dias de Amorim, Farid Benbadis, Vania Conan, and John Whitbeck. The Accordion Phenomenon: Analysis, Characterization, and Impact on DTN Routing. In *Proceedings of the 28rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, pages 1116–1124. IEEE, 2009.

[52] Tomasz Tylenda, Ralitsa Angelova, and Srikanta Bedathur. Towards time-aware link prediction in evolving social networks. In *Proceedings of the 3rd Workshop on Social Network Mining and Analysis (SNA-KDD '09)*, 2009.

[53] A. Vázquez, R. Dobrin, D. Sergi, J. P. Eckmann, Z. N. Oltvai, and A. L. Barabási. The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *PNAS*, 101:17940–17945, 2004.

[54] A. Vázquez, J.G. Oliveira, and A.-L. Barabasi. The inhomogeneous evolution of subgraphs and cycles in complex networks. *Physical Review E*, 71:025103, 2005.

[55] Chao Wang, Venu Satuluri, and Srinivasan Parthasarathy. Local probabilistic models for link prediction. In *Data Mining, IEEE International Conference on*. IEEE Computer Society, 2007.

[56] Xiaoming Wang, Zhongmei Yao, and Dmitri Loguinov. Residual-based measurement of peer and link lifetimes in gnutella networks. In *INFOCOM*, pages 391–399, 2007.

[57] Walter Willinger, David Alderson, and Lun Li. A pragmatic approach to dealing with high-variability in network measurements. In *Internet Measurement Conference*, pages 88–100, 2004.