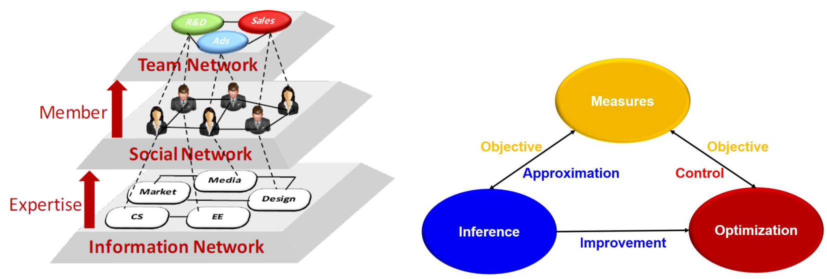
Network Connectivity in Complex Networks: Measures, Inference and Optimization

Chen Chen and Hanghang Tong

Arizona State University, Computer Science Department
Tempe, Arizona, 85281, United States
{chen_chen, hanghang.tong}@asu.edu

1 Problem Description

Networks naturally appear in many high-impact applications, ranging from epidemic study, social network mining to infrastructure analysis. As the world is becoming increasingly connected and coupled, nodes from different application domains tend to be inter-dependent on each other, forming the so-called multi-layered networks [3]. One classic example of multi-layered network is collaboration platform as shown in Fig. 1(a), where team collaboration network is supported by the social network between employees, which is backed by the information network among knowledge bases. Compared to single-layered networks, multi-layered networks are more vulnerable to external attacks since a small disturbance in one layer may cause catastrophic failures of the entire system through dependency links. By combining the dependency links with single networks from different domains, we aim to derive a generic analytic framework that can be applied to both single-layered networks and multi-layered networks.



(a) Multi-layered collaboration network (b) Overview of the problem

Fig. 1: Example of multi-layered network and problem overview.

Among the various network properties in the literature, network connectivity is the one that shows close correlation to network robustness and dissemination ability. Some prevalent connectivity measures include natural connectivity [8], epidemic threshold [1], etc. In real applications, it is of key importance to optimize (maximize/minimize) the connectivity of the network by manipulating its underlying structure within certain budget. As the connectivity measures varies from one to the other, different optimization algorithms are designed to optimize each of them (e.g. [5] for epidemic threshold and [2] for natural connectivity). Although there is no panacea for the various connectivity optimization problems, those specialized optimization methods still share many properties in common, which inspires us to seek for a general optimization framework that could fit into a wide range of connectivity measures.

On the other hand, most of the above-mentioned works assume that the network structure is accurate and static, which is not realistic in many applications. The observed links in the network may be mistakenly placed due to noise, while true connections might be missed due to incomplete data sources. To address this problem, it is important to infer a more accurate network structure from the observed links. Furthermore, as the network is evolving over time, its key properties would vary from time to time. To keep track of those changing properties in a timely manner, it is necessary to have a fast inference algorithm to provide good estimations on the corresponding properties.

To summarize, the main problems studied in our work are focused on measures, inference and optimization of network connectivity in complex networks. The relationship between those problems are shown in Fig. 1(b). Generally speaking, a well defined connectivity measure serves as the objective to inference and optimization tasks; The inference results in turn provide a good approximation on the connectivity measure and improve the accuracy of the input network for optimization tasks; Last, the optimization methods are used to find optimal strategies to manipulate the network structure, which can effectively change the connectivity of the network.

2 Methodologies

In this section, we present our approaches for the three problems.

Measures. We unify a family of prevalent network connectivity measures (SUBLINE) by viewing the connectivity of the entire network as an aggregation over the connectivity scores of its sub-networks (e.g. subgraphs, motifs) [3]. Based on this definition, we find that a variety of connectivity measures can be viewed as instances of the SUBLINE model by carefully tuning the model coefficients (e.g. epidemic threshold, natural connectivity, etc.).

Inference. To infer a complete set of dependency links in multi-layered networks, we propose a collective collaborative filtering based method that could effectively identify missing links across the layers. We have compared our algorithm with a wide range of baseline methods under various evaluation metrics (MPR, HLU, MAP and AUC) [6]. Experiment results show that our algorithm outperforms all other baselines in all the metrics. On the other hand, in dynamic network settings, we derive an efficient scheme to approximate the algebraic connectivity measures by leveraging matrix perturbation theory [4]. Experiment results demonstrate that our methods can effectively keep track of a wide range of connectivity measures with up to $20\times$ speedup in a fairly long period of time.

Optimization. Based on the definition of SUBLINE model, we find that for any SUBLINE connectivity measures, the corresponding connectivity control problem enjoys the diminishing returns property in both single-layered networks and multi-layered networks, which naturally lends itself to a family of provable near-optimal control algorithms with linear complexity [3]. We have compared our optimization algorithms with other heuristic baseline methods on their abilities of minimizing network connectivity measures under certain budget. Experiment results show that our algorithm outperforms all other baselines in all optimization scenarios.

The effectiveness of our proposed methods is validated on various datasets from three different domains, ranging from biology, infrastructure to social collaboration. In the biology domain, we have a three-layered network system that depicts the inter-dependency between proteins, genes and drugs [7]. While in the infrastructure domain, we construct a three-layered infrastructure system that shows the interactions between power grid, router network and transportation network [6]. In social collaboration domain, we build a three-layered academia collaboration platform from Aminer dataset [9], which reveals the connections between researchers, papers and conference venues.

3 Future Work

Our future research aims to dig deeper into the three problems on network connectivity optimization. From the connectivity measures aspect, we want to explore some local connectivity measures can be used to evaluate the connectivity based centrality for individual node or a small group of nodes; For the inference problem, we would like to derive a more general network completion method that can jointly identify both within-layer missing links and cross-layer missing dependencies via network embedding methods. Last, for the connectivity optimization part, our goal is to design a more effective optimization scheme that can further improve the lower bound of current approximation algorithms.

References

1. Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., Faloutsos, C.: Epidemic thresholds in real networks. ACM Transactions on Information and System Security (TISSEC) 10(4), 1 (2008)
2. Chan, H., Akoglu, L., Tong, H.: Make it or break it: Manipulating robustness in large networks. In: Proceedings of the 2014 SIAM International Conference on Data Mining. pp. 325–333. SIAM (2014)
3. Chen, C., He, J., Bliss, N., Tong, H.: On the connectivity of multi-layered networks: Models, measures and optimal control. In: 2015 IEEE International Conference on Data Mining, ICDM 2015, Atlantic City, NJ, USA, November 14-17, 2015. pp. 715–720 (2015)
4. Chen, C., Tong, H.: Fast eigen-functions tracking on dynamic graphs. In: Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, BC, Canada, April 30 - May 2, 2015. pp. 559–567, ***Bests of SDM 2015*** (2015)
5. Chen, C., Tong, H., Prakash, B.A., Tsourakakis, C.E., Eliassi-Rad, T., Faloutsos, C., Chau, D.H.: Node immunization on large graphs: Theory and algorithms. IEEE Trans. Knowl. Data Eng. 28(1), 113–126 (2016)
6. Chen, C., Tong, H., Xie, L., Ying, L., He, Q.: FASCINATE: fast cross-layer dependency inference on multi-layered networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016. pp. 765–774, ***Bests of KDD 2016*** (2016)
7. Galperin, M.Y., Rigden, D.J., Fernández-Suárez, X.M.: The 2015 nucleic acids research database issue and molecular biology database collection. Nucleic acids research 43(D1), D1–D5 (2015)
8. Jun, W., Barahona, M., Yue-Jin, T., Hong-Zhong, D.: Natural connectivity of complex networks. Chinese physics letters 27(7), 078902 (2010)
9. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: Arnetminer: extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 990–998. ACM (2008)