Research Statement

Sungsu Lim

March 1, 2018

My research focuses on modeling and mining large-scale datasets, with an emphasis on social and information networks. Various kinds of complex systems can be represented as networks. In the era of big data, massive amount of data are being generated everyday, which leads to the importance of analyzing large-scale networks. The goal of my research is to understand structural properties and dynamical processes in complex networks. I am particularly interested in developing new methods and algorithms to solve real-world problems with theoretical guarantee and various applications.

1 Research Experience

During my doctoral studies, I have worked on the following problems:

- **Detecting Community Structures in Social Networks**: Community detection is a fundamental problem in social network analysis. Many models and methods have been developed for this purpose. However, owing to a wide variety of network structures, there remain challenges to determining community structures from social networks. My research focused on four important problems: identifying the community structure of a network when it consists of (i) overlapping community structure, (ii) weak strength community structure, (iii) complex substructure, and (iv) multi-layer structure. My research proposed a transformation-based algorithm that converts an original graph to a transformed graph that reflects the structure of the original network and has a superior structure for each problem.

  For the first problem, we proposed the link-space transformation. This transformation provides a weighted graph that is easier to find link communities that can be converted to the vertex communities whose memberships can overlap. Based on the link-space transformation, we developed an overlapping community detection algorithm $LinkSCAN^*$ [1]. The proposed algorithm outperforms existing algorithms, especially for networks with many highly overlapping nodes.

  For the second problem, we proposed the BlackHole transformation. This transformation is a force-directed graph embedding that attempts to place the vertices of the same community at a single position. The algorithm $BlackHole$ maps the graph into a space using the transformation and then group vertices by a conventional clustering algorithm [2]. The proposed algorithm achieves high performance regardless of the difficulty of community detection in terms of the mixing.

  For the third problem, we proposed a motif-based embedding method for graph clustering by modeling higher-order relationships among vertices in a graph [3]. The proposed method considers motif-based attractive forces to enhance the clustering tendency of points in the output space of graph embedding. We showed that the motif-based embedding is more effective in detecting communities than existing graph embedding methods both theoretically and experimentally.

  For the fourth problem, we proposed a framework of differential flattening, which facilitates the analysis of multi-layer graphs, and applied this framework to community detection [4]. Differential flattening merges multiple graphs into a single graph such that the graph structure with the maximum clustering coefficient is obtained from the single graph. We developed an algorithm of finding the optimal merger by solving a nonlinear optimization problem.

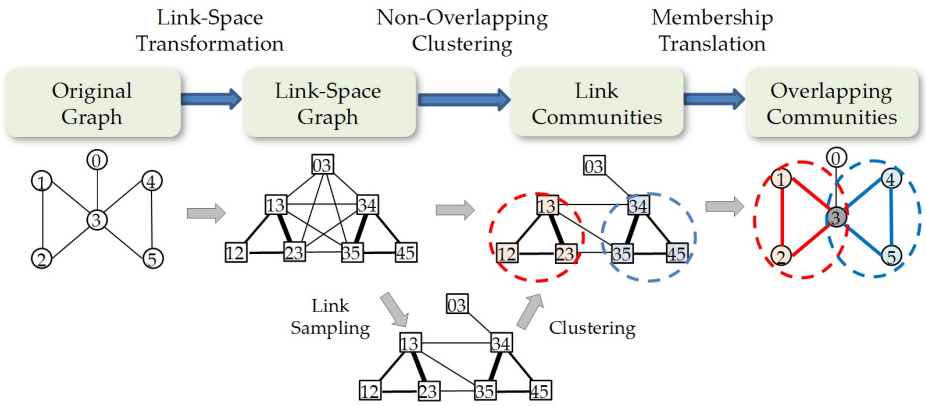

Figure 1: The overall procedure of *LinkSCAN** [1].

- **Analyzing Information Diffusion in Complex Networks**: Information diffusion plays an essential role in numerous human interactions, including diffusion of disease and rumors. Understanding how information flow on networks is a central problem for industry and academia. We developed iterative methods for estimating the spread of information in networks with various structures under the general threshold model [5] and the independent cascade model [6]. Using the fixed point theorem of nonlinear equation, we analyzed the tipping points of the models. Concentration inequalities assure that our methods approximate the sizes of cascade well. We also designed the behavior and dynamics of competing sources by proposing a generalized information diffusion model of having multiple states [7, 8]. We used a generalized mean-field approximation to analyze the proposed model, and showed that our methodology predicts the diffusion accurately under variety of graphs.

  In addition, we analyzed the network dynamics in communication networks. Max-Weight algorithm has been one of the most studied algorithm for routing and scheduling in wireless networks. It keeps the queues in network stable, *i.e.*, keeps the queue sizes bounded, for a fixed set of flows. We defined a general adversarial model in which any interference conditions are possible and edge rates can vary over time. This allows us to capture arbitrary types of wireless interference behavior. Under the model, we proved that the Max-Weight keeps all the queues stable whenever this is feasible by some routing and scheduling algorithm. [9, 10].

2 Future Work

I am currently working on the following challenging problems:

- **Developing Network Analysis using Higher-Order Relationships**: A notable trend in graph clustering is to consider higher-order graph substructures, called the network motifs or graphlets. The order of a graph substructure is determined by the number of vertices required to express the substructure. Previous studies on graph clustering have mainly focused on considering pairwise relationships between vertices, *i.e.*, second-order substructures. It is known that using higher-order relationships that involve multiple vertices can improve the understanding of the underlying graph structure. For instance, a triangle, which is a third-order relationship between vertices, is regarded as a fundamental substructure for understanding social networks and their communities.

- **Detecting Complex Patterns with Sensor Networks**: Another goal of my research is to detect complex patterns using sensor network datasets, with good accuracy. Recently, there has been an explosion of interest in mining time series and trajectory data. Extracting interesting features from the streaming sensor data is a challenging problem. I am planing to design new models to analyze and predict the behavior patterns in complex social processes. I believe that developing complex network analysis and big data analytics can be a key solution for the research problem.

3 Conclusion

My research aims to build efficient frameworks to understand the large-scale network systems. As a researcher of data mining and network science, I conduct research to analyze complex networks on massive datasets both in theoretical and applied perspectives. Toward this goal, I intend to extend the expertise of mathematical and analytical tools, as well as machine learning techniques. I look forward to working closely with researchers in multidisciplinary areas.

References

[1] S. Lim, S Ryu, S. Kwon, K. Jung, and J.-G. Lee. LinkSCAN*: Overlapping community detection using the link-space transformation. In *Proc. of the 30th IEEE ICDE*, pages 292–303, 2014.

[2] S. Lim, J. Kim, and J.-G. Lee. BlackHole: Robust community detection inspired by graph drawing. In *Proc. of the 32nd IEEE ICDE*, pages 25–36, 2016.

[3] S. Lim and J.-G. Lee. Motif-based embedding for graph clustering. *Journal of Statistical Mechanics: Theory and Experiment*, 2016:123401, 2016.

[4] J. Kim, J.-G. Lee, and S. Lim. Differential flattening: A novel framework for community detection in multi-layer graphs. *ACM Trans. on Intelligent Systems and Technology*, 8(2):27, 2017.

[5] S. Lim, I. Jung, S. Lee, and K. Jung. Analysis of information diffusion for threshold models on arbitrary networks. *European Physical Journal B*, 88(8):201, 2015.

[6] S. Lim, J. Shin, N. Kwak, and K. Jung. Phase transitions for information diffusion in random clustered networks. *European Physical Journal B*, 89(9):188, 2016.

[7] S. Lim, K. Jung, and J. C. S. Lui. Phase transition of multi-state diffusion process in networks. *ACM SIGMETRICS Performance Evaluation Review*, 41(3):31–33, 2013.

[8] Y. Lin, J. C. S. Lui, K. Jung, and S. Lim. Modeling multi-state diffusion process in complex networks: Theory and applications. *Journal of Complex Networks*, 2(4):431–459, 2014.

[9] S. Lim, K. Jung, and M. Andrews. Stability of the max-weight protocol in adversarial wireless networks. In *Proc. of the 31st IEEE INFOCOM*, pages 1251–1259, 2012.

[10] S. Lim, K. Jung, and M. Andrews. Stability of the max-weight protocol in adversarial wireless networks. *IEEE/ACM Trans. on Networking*, 22(6):1859–1872, 2014.