MAE-253: Homework 2a

John Karasinski

May 5, 2016

Project Overview

Project Members:

Ehsan Gholami, John Karasinski, Jenette Sellin, Dmitry Shemetov

We are looking to perform community detection in online information sharing communities, and to compare network properties across different communities. We have large data sets from Reddit, StackExchange, and HackerNews, and we're looking to build networks from the commenting dynamics contained within. One of the more interesting phenomena in online communities is Eternal September. Simply put, this is when the popularity of a community reaches a critical threshold beyond which no central core of the members are in control; new users constantly come in and bring the signal to noise ratio down heavily.

Literature Review

Fast unfolding of communities in large networks [1]

This paper proposes the "Louvain" method, a community detection technique which they show to outperform prior techniques in terms of computation time. The technique attempts to maximize the modularity of the groups by a two step process. The technique first locally maximizes the modularity by changing the community of each node, then aggregates the nodes and links in each community to create a new network of communities. This process iterates until the modularity is maximized. They go on to validate their method with a Belgian mobile phone network, which they also show to be substantially faster then several other methods.

[1] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008. http://arxiv.org/pdf/0803.0476.pdf

Community structure in social and biological networks [2]

This paper introduces a new method of detecting community structure by focusing on eges which the least central between communities. These least central edges are the most between communities, and communities are detected by iteratively removing edges from the original graph. This algorithm has a worst-case run time on the order of m^2n, where m is the number of edges, and n is the number of vertices, though the authors claim that the actual run time is much better for networks with large community structure. The authors go on to test their method on two real-world networks and

[2] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12): 7821–7826, 2002. http://arxiv.org/pdf/cond-mat/0112110.pdf

show that there is good agreement between their results and previous results from other methods.

Network motifs: simple building blocks of complex networks. [3]

This paper studies the network motifs present in complex networks, and investigate those that occur much more frequently than one would expect from a random network. They define motifs as "recurring, significant patterns of interconnections", and create an algorithm to detect motifs with node size $n = 3$ and $n = 4$. The authors apply their algorithm to several real world networks, and make observations about the motifs that appear frequently among them. They finish by saying that these motifs are present at all network and subnetwork sizes, making the motifs a good identifier of the network or subnetwork.

[3] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594): 824–827, 2002. http://www.sns.ias.edu/~tlusty/courses/InfoInBio/Papers/AlonMotifs2002.pdf

Expertise networks in online communities: structure and algorithms [4]

This paper studies an online community, the Java Forum, and compares different algorithms to measure user expertise. They characterize the network as having a bow tie structure, similar to how others have characterized the web. The authors go on to calculate the expertise of users based off simple statistical measures (ie, how many posts a user makes), z-score measures, a page-rank-esque algorithm, and HITS Authority. The authors went on to have two 'Java programming experts' manually rank a random sampling of 135 users by their expertise on a 5 level scale, from 'Newbie' to 'Top Java expert'. They then compare the results of all of these algorithms, and find that the simple metrics perform as-well or better than the more complicated methods, and that structural information of the networks can be used to evaulate the expertise of individual users. The authors finish by using these results to model and simulate the Java Forum network to study the dynamics of the communities within.

[4] Jun Zhang, Mark S Ackerman, and Lada Adamic. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*, pages 221–230. ACM, 2007. http://www.wwwconference.org/www2007/papers/paper516.pdf

References

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008. http://arxiv.org/pdf/0803.0476.pdf.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002. http://arxiv.org/pdf/cond-mat/0112110.pdf.

Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002. http://www.sns.ias.edu/~tlusty/courses/InfoInBio/Papers/AlonMotifs2002.pdf.

Jun Zhang, Mark S Ackerman, and Lada Adamic. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*, pages 221–230. ACM, 2007. http://www.wwwconference.org/www2007/papers/paper516.pdf.