Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference

Hierarchical Fuzzy Spectral Clustering in Social Networks Using Spectral Characterization

Scott Wahl* and **John Sheppard**
Department of Computer Science
Montana State University
Bozeman, MT 59717

Abstract

An important aspect of community analysis is not only determining the communities within the network, but also sub-communities and hierarchies. We present an approach for finding hierarchies in social networks that uses work from random matrix theory to estimate the number of clusters. The method analyzes the spectral fingerprint of the network to determine the level of hierarchy in the network. Using this information to inform the choice of clusters, the network is broken into successively smaller communities that are attached to their parents via Jaccard similarity. The efficacy of the approach is examined on two well known real world social networks as well as a political social network derived from campaign finance data.

Introduction

Many real world networks are characterized by dense sub-networks that are commonly referred to as communities and are generally composed of groups of nodes that have elements in common with each other. Examples of networks that have community structure can be drawn from social (Fortunato 2009), biological (Power et al. 2011), gene expression (Zhang and Horvath 2005), and many other types of networks. Since the communities can represent fundamental properties of the network, their discovery is important for understanding the nature of the networks (Newman and Girvan 2004),(Flake et al. 2002). The primary focus of this paper is on social networks.

To best represent the communities, a classification of the nodes into clusters should satisfy two important realities of many social networks: overlap and hierarchy. For the first, nodes within the network may belong to multiple communities. Much like in human social groups, an individual may belong to more than one community or have multiple affiliations (Zhang, Wang, and Zhang 2007). Hierarchy is another important aspect of some social networks wherein smaller communities together make up larger ones. Military, business, and political hierarchies are all examples where individual smaller groups combine into a larger group.

*also affiliated with the National Institute on Money in State Politics

There is already a wealth of research on finding communities within networks. Some initial work focused on crisp splits of the network into non-overlapping, non-hierarchical communities (Newman and Girvan 2004), (Newman 2006). As part of this, a method for evaluating the quality of a partitioning of the data into clusters was developed called modularity. The idea behind modularity is to determine how well a community split describes the likelihood of the community as it relates to a null model, where each node keeps the same degree but is connected at random to other nodes. This is defined by

$$Q = \sum_{i \in C} \left(e_i - a_i^2 \right)$$

where C is the set of communities, e_i is the fraction of edges between the nodes in community i, and a_i is the fraction of edges that connect to nodes in community i, regardless of source.

A variety of other approaches have been developed for finding communities in networks (Pons and Latapy 2004), (Blondel et al. 2008). A very popular method is spectral clustering (Pothen, Simon, and Liou 1990), (Ng, Jordan, and Weiss 2001). These have proved popular for their ease of implementation and their ability to handle non-convex clusters.

While some of the above mentioned methods can yield hierarchies, they do not find overlapping communities. To find these fuzzy communities, a variety of approaches have been presented. Palla uses a clique percolation method to find adjacent cliques with overlapping nodes (Palla et al. 2005). Other methods use fuzzy modularity and simulated annealing or other techniques to find relevant partitions (Liu 2010), (Bandyopadhyay 2005), (Xie, Szymanski, and Liu 2011). Fuzzy c-means is another possibility for determining fuzzy clusters and has been used to find hierarchies of clusters (Torra 2005), (Devillez, Billaudel, and Lecolier 2002). The approach presented here differs in its use of spectral clustering and spectral characterization to create a top-down algorithm for finding hierarchical fuzzy clusters.

In addition to the above mentioned methods, there has also been work in social networks that change over time and methods for tracking and predicting communities (Hopcroft et al. 2004), (Spiliopoulou et al. 2006), (Aynaud and Guillaume 2010). One such method attempts to predict the emergence of future communities using link prediction (Jung and Segev 2014).

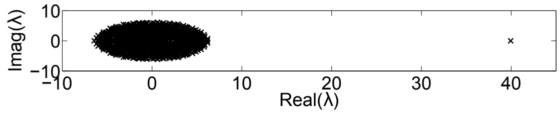

Figure 1: Random Network

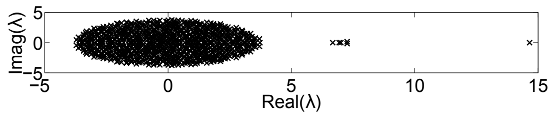

Figure 2: Random Network with Communities

Related Work

A common issue with determining the number of clusters is figuring out an appropriate number of clusters to begin the process, or an appropriate and fitting place to stop joining networks in the case of agglomerative methods. To find the number of clusters, another possibility involves analyzing the eigenvalues of the adjacency matrix.

Spectral Characterization

To find the optimal number of clusters in the network, the top down spectral approach defined later uses properties of the eigen-spectrum of the adjacency matrix. Prior work has shown that the eigen-spectrum of a network can reveal certain properties of the community structure of that network. The Perron-Frobenius theorem for non-negative matrices indicates that the largest magnitude eigenvalue is real and positive (MacCluer 2000). Figure 1 shows the eigenvalues of a randomly created directed network with no community structure. As predicted by previous studies, the largest eigenvalue is well outside the primary cloud formed by the other eigenvalues. No other large eigenvalues were expected due to the uniformly distributed connections. Undirected networks have similar form, but have only real valued eigenvalues. The network in the first example has 1000 nodes and an average combined in/out degree on those nodes of 40.

Based on other work, in general, a network with k communities will have k large eigenvalues (Chauhan, Girvan, and Ott 2009), (Sarkar and Dong 2011), (Sarkar, Henderson, and Robinson 2013). As an example, Figure 2 shows a network created with 6 communities and the 6 large eigenvalues from its spectrum. There are now 6 points outside the cloud, corresponding to the communities.

In the case of networks with hierarchical communities, the spectrum of the network shows multiple groups of eigenvalues when the communities are of similar size. Figure 3 has a top level hierarchy of 4 nodes with each having 4 subcommunities, creating 16 total clusters. As can be seen in the graph, 16 eigenvalues are located outside of the main cloud and are split into two separate clusters. The gap between eigenvalues, or eigen-gap, indicates a separation between levels within the hierarchy. This principle is what is used when attempting to determine the number of communities in each hierarchical level relevant for spectral clustering.

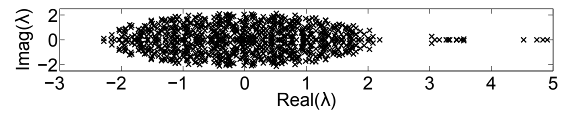

Figure 3: Random Hierarchical Network

Fuzzy Modularity

Since the base equation for modularity requires crisp communities, a fuzzy modularity $\widetilde{Q}$ was developed to assess the splits created by fuzzy spectral clustering (Zhang, Wang, and Zhang 2007). Its form and principle are similar to that of the original modularity. Given an adjacency matrix A where $A_{ij} \neq 0$ indicates an edge between i and j, $\widetilde{Q}$ is defined by

$$\widetilde{Q}\left(U_k\right) = \sum_{c=1}^{k} \left[\frac{E\left(\bar{V}_c, \bar{V}_c\right)}{E\left(V, V\right)} - \left(\frac{E\left(\bar{V}_c, V\right)}{E\left(V, V\right)} \right)^2 \right].$$

In the above equation, U_k is a fuzzy partition of k clusters,

$$E\left(\bar{V}_c, \bar{V}_c\right) = \sum_{i \in \bar{V}_c, j \in \bar{V}_c} A_{ij} \left(\frac{u_{ic} + u_{jc}}{2} \right),$$

$$E\left(\bar{V}_c, V\right) = E\left(\bar{V}_c, \bar{V}_c\right) + \sum_{i \in \bar{V}_c, j \in V \setminus \bar{V}_c} A_{ij} \left(\frac{u_{ic} + u_{jc}}{2} \right),$$

and

$$E\left(V, V\right) = \sum_{i \in V, j \in V} A_{ij}$$

where u_{ic} is the fuzzy assignment of node i to cluster c. For a node i to be a part of cluster $\bar{V}_c$, the fuzzy value assigned for that cluster must exceed a certain threshold: $\bar{V}_c = \{i | u_{ic} > \lambda, i \in V\}$.

By using fuzzy modularity, it is possible to perform similar agglomerative clustering techniques as those used for regular modularity (Havens et al. 2013), or other approaches, like simulated annealing (Liu 2010).

Approach

The approach proposed here is primarily based on the spectral clustering work of Ng, Jordan, and Weiss (Ng, Jordan, and Weiss 2001) as well as Zhang, Wang, and Zhang (Zhang, Wang, and Zhang 2007).

First, the spectral composition of the network must be determined. As described in prior work, this aids in determining the level of hierarchy in the network and the number of clusters at each hierarchical level. If the hierarchical structure is weak, it is possible to fall back on iterative testing of the partitions using fuzzy modularity as an optimization metric.

With the number of clusters at the base level determined, an initial clustering is performed on the square adjacency matrix A using the following technique.

- Let D be a diagonal matrix where $D_{i,i}$ is the sum of the i-th row of A. This is equivalent to the weighted degree of each node.

- Construct the Laplacian matrix $L = D^{-1/2}AD^{-1/2}$.
- Determine the k largest eigenvectors, $x_1, x_2, ..., x_k$ of the Laplacian L and create the matrix $X = (x_1, x_2, ..., x_k]$. X is then normalized such that each row has unit length.
- Using X, perform fuzzy c-means clustering on the data to obtain U, a $n \times k$ matrix where k is the number of clusters and n is the number of data points in A.

This procedure obtains the top level communities of the network. To obtain hierarchical structure, the process is repeated with a varying k corresponding to the number of clusters in each hierarchical level. Each level is connected to its previous by calculating the Jaccard similarity measure of the communities.

$$\mathcal{J}\left(\bar{V}_i, \bar{V}_j\right) = \frac{\left|\bar{V}_i \cap \bar{V}_j\right|}{\left|\bar{V}_i \cup \bar{V}_j\right|}$$

The results give similarity measures for the smaller clusters that can be used to assign each cluster to its best matching parent.

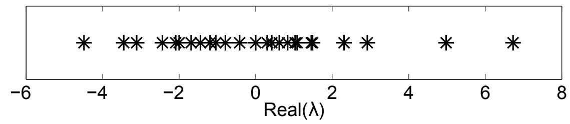

Figure 4: Karate Spectral Characteristic

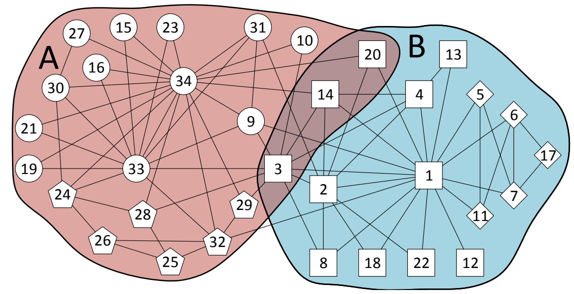

Figure 5: Karate network overlapping communities: $k = 2$ and $\lambda = .25$

Experiment

To test the efficacy of the algorithm, we analyze three real world networks and present the fuzzy clustering results for those networks. Two of these networks were chosen for their popularity in benchmark testing for community detection as well as the presence of hierarchical communities. The third is a new network derived from campaign finance data.

For each data set, the set of eigenvalues is examined to determine the hierarchical structure of the network. Using this information, the network is partitioned into each of the different levels of communities using fuzzy spectral clustering. Each level is attached to its parent via the best Jaccard similarity measure for that child and parent.

Zachary Karate Network

The first example is the Zachary Karate Club network (Zachary 1977), a common benchmark used with community detection algorithms. It is popular since it is small and has known clusters. Due to conflict between the club president and the instructor, the 34 members split into two separate groups. This network has another useful property in that there are sub-communities within the two primary groups. The best partition of the network, with respect to modularity, splits the set into four groups (Fortunato 2009). This network and the known clusters and sub-clusters are shown in later figures. The node shapes represent the true best partitioning of the network by modularity and are there for reference. For the shapes, the circles and pentagons combine into one true cluster, and the squares and diamonds combine to create the other.

The corresponding spectrum for the network is given in Figure 4. This spectrum shows two hierarchical levels, based on the large gaps between eigenvalues located outside the cloud. The two largest eigenvalues correspond to the communities created by the true clusters. Outside of the primary cloud is another cluster of eigenvalues that represent the subcommunities within the primary clusters. Using this information, hierarchical fuzzy spectral clustering is applied to the network.

Using fuzzy spectral clustering as defined earlier, Figure 5 shows the overlapping clusters with $k = 2$. Assigning communities with $\lambda = 0.25$, nodes 3, 14, and 20 are considered to be overlapping nodes. This appears to make sense as those nodes are connected to the most connected and central nodes of the two different clusters. For cluster $\mathbf{A}$, these are nodes 1 and 2, while in $\mathbf{B}$ these are 33 and 34.

Now, these results are compared with the sub-clusters. Figure 6 shows the fuzzy clusters on $k = 4$ and $\lambda = 0.16$. At this level, 3, 14, and 20 are no longer overlapping nodes due to the dissimilarity of the clusters. These clusters are now less defined by their proximity to the central nodes 1, 2, 33, and 34, and instead more by their local connections. Thus, the set $\{1, 5, 6, 7, 11, 17\}$ becomes its own cluster since most of these nodes are only connected to each other. The set $\mathbf{A_2} = \{24, 25, 26, 28, 29, 32\}$ in now its own community, separate from $\mathbf{A_1}$, which are better defined by their proximity to 33 and 34. Additionally, nodes 9, 10, and 31 become assigned to both $\mathbf{A_1}$ and $\mathbf{B_1}$.

It should be noted that increasing the value of λ to $\lambda = 0.3$ results in an assignment with no overlap where the communities are identical to the original sub-communities.

Dolphin Network

The Dolphin network is another well known example of a social network (Lusseau and Newman 2004). This network represents a group of dolphins that had been tracked over a period of time. Eventually, the dolphins split into separate groups. From prior work by Lusseau and Newman, one of the communities was further broken down into smaller communities. In the later figures, the circular nodes represent one of the true clusters and all of the other shapes together form

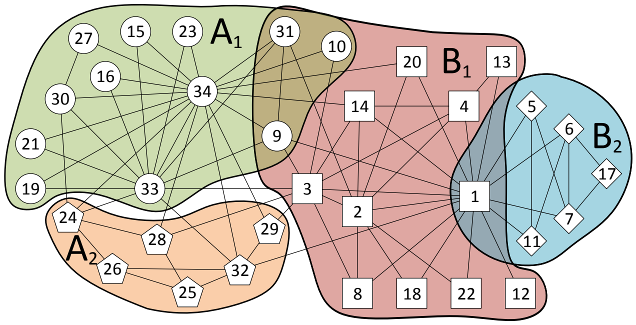

Figure 6: Karate network overlapping communities: $k = 4$ and $\lambda = .16$

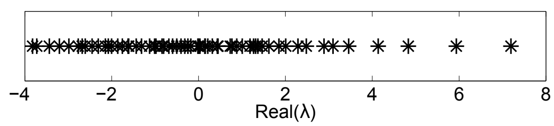

Figure 7: Dolphin Spectral Characteristic

the other true cluster. The individual shapes correspond to the sub-cluster results from (Lusseau and Newman 2004).

Viewing the spectral characteristics of this network in Figure 7, it is possible to see by that standard that it does not have as strong of a hierarchical nature when compared to the karate network. There are two hierarchical levels, but the exact number of sub-communities is difficult to determine as it begins to merge with the primary cloud.

As the largest eigen-gap between the values occurs after the second largest eigenvalue, the initial pass clusters using two partitions. The next phase proves more difficult due to the remaining eigenvalues. Since there is a fairly smooth transition from the bulk distribution to the other eigenvalues, we calculate the best fuzzy modularity for each possible partitioning, restricting the search to the approximate number of communities. This procedure gives a best partition using six clusters.

Using this information to get the smaller clusters, the resulting six communities are given in Figure 9. These communities align well with the previous results, with the exception of 42 and 43, which are added to the community $\mathbf{A_2}$. Unfortunately, these two have considerably different connections in relation to the rest of the members of $\mathbf{B_3}$, weakening its association with those nodes. Since the fuzzy assignment across all nodes must equal 1, this gets distributed across the other nodes, raising the association with $\mathbf{A_2}$ beyond the λ threshold. Likely, it is most strongly tied with these because $\mathbf{A_2}$ and $\mathbf{B_3}$ share proximity to $\mathbf{B_4}$. Raising λ does place them solely in $\mathbf{B_3}$, but it weakens other associations and yields lower modularity.

Still, even with that outlier, communities $\mathbf{B_k}$ closely correspond to one of the true communities. Likewise, communities $\mathbf{A_k}$ match closely with the other true community.

Although there are now more communities than what was

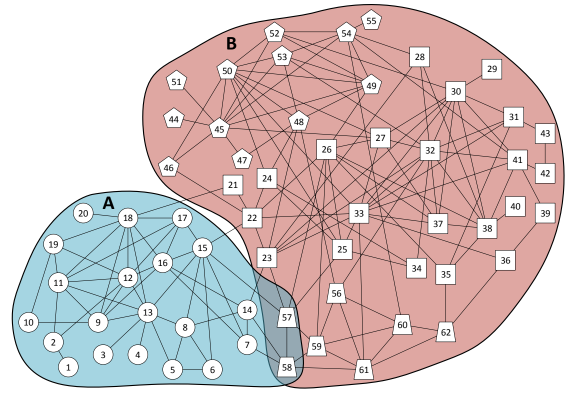

Figure 8: Dolphin network overlapping communities: $k = 2$ and $\lambda = .20$

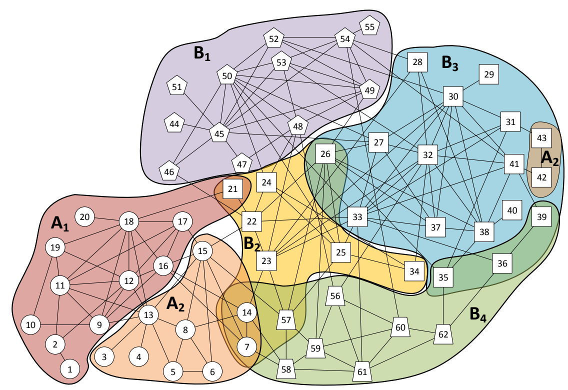

Figure 9: Dolphin network overlapping communities: $k = 6$ and $\lambda = .17$

determined by Lusseau and Newman, by merging $\mathbf{B_2}$ and $\mathbf{B_3}$ into a community, and $\mathbf{A_1}$ and $\mathbf{A_2}$ into another community, the results are very similar. Attempting to directly compute $k = 4$ communities yields different results as shown in Figure 10.

Alaska Campaign Finance

This data set is not one heavily analyzed by social network algorithms and is obtained from the National Institute on Money in State Politics (NIMSP)[1]. NIMSP gathers data on campaign finance from state and federal elections. Previous work has shown that the primary motivator for donations from individuals is ideology. However, for non-individuals, they may attempt more strategic donations (Bonica 2014). This premise is tested here with hierarchical fuzzy spectral clustering. This particular set is compiled from donations that were reported in Alaska during 2012 for general elections and represents business and other non-individual donations to candidates, creating a bipartite network where each node is a candidate or donor and an edge is a donation.

In preparing the data set, nodes were removed if the node

[1]http://www.followthemoney.org

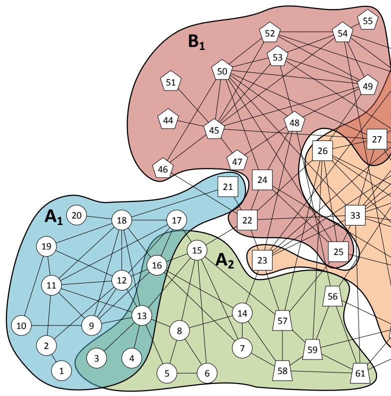

Figure 10: Dolphin network overlapping communities: $k = 4$ and $\lambda = .27$

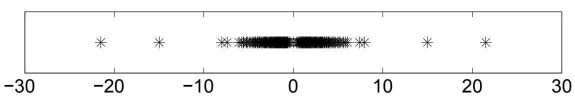

Figure 11: Alaska Spectral Characteristic

only gave once over the course of the election cycle. Similarly, candidates with only one donor were also removed. Multiple donations from a donor to the same candidate were also removed. The remaining data covers 214 nodes and 1426 edges. Despite the simplicity of the graph here, the full scope of information available for creating a fully featured and heterogeneous network from campaign finance is substantial and growing rapidly. Figure 11 shows the spectral characteristic of this network. It is rather similar in nature to the Zachary karate network in that it indicates two clusters at the top hierarchy and four clusters at a second hierarchical level. The large negative eigenvalues on the left of the graph are due to the graph being bipartite and can be ignored.

After obtaining the hierarchy from the above method, there are two communities at the top level, each of which have two child communities. Unsurprisingly, in the parent communities $\mathbf{D}$ and $\mathbf{R}$, the candidates mostly split on party lines, and ideology appears to dominate donations. The overlap between the two groups is especially interesting, however, as it includes donors who gave evenly between Democrats and Republicans in Alaska. Moreover, the candidates in this overlap were overwhelmingly winners, with only one candidate losing the election.

To verify these results, the overlapping donors were checked against their entire historical data record. Based on this data from NIMSP, the overlap donors have on average given much more to winning candidates with very little variation. The rest of the donors have not done as well at donating solely to winners, though there is far more variation in the percentage of dollars that went to winning candidates (Table 1).

Analyzing the sub-communities at the next hierarchy, there is a clear pattern in the candidates within each community. Analyzing each community separately,

Nodes	Ratio to Winners	Std. Dev. of %
In Both $\mathbf{D}$ and $\mathbf{R}$	4.94	0.094
Solely in $\mathbf{D}$	1.20	0.232
Solely in $\mathbf{R}$	2.80	0.182

Table 1: Historical Alaska Donations

- $\mathbf{D_1}$ comprises Democratic candidates exclusively, 83% of which lost the election. The donors have given to Democrats with only one donation ever to a Republican candidate and one to an unaffiliated candidate.
- $\mathbf{D_2}$ has mostly Democratic candidates at 88%, 56% of which won their election. The donors have given almost four times as much to Democrats as Republicans.
- $\mathbf{R_1}$ comprises 10 Democratic and 28 Republican candidates. These candidates were almost exclusively winners, with only one losing. Similar to $\mathbf{D_2}$, the donors gave four times as much, in this case favoring Republicans.
- $\mathbf{R_2}$ contains only Republican candidates as well as a single unaffiliated candidate. Only 55% of these candidates won the election. Over the years the donors in this group have given over 54 times as much to Republicans as Democrats.

For the children of $\mathbf{D}$, those who gave exclusively to Democrats generally gave to losing candidates while those who gave more evenly donated more to winners. Regarding the children of $\mathbf{R}$, while the donors who gave exclusively to Republicans chose more winners, those who gave to Democrats as well picked almost nothing but winning candidates. This shift may be due to the overall political leanings of Alaska where their legislature has a majority of Republicans.

Conclusion

As shown by the previous experiments, the given fuzzy hierarchical clustering method performs well on many real world data sets. One limitation of the test networks used is that they do not have deep hierarchies. All of the networks have only two levels that can be seen by the spectral characterization. The efficacy of the algorithm on real world data sets with even more levels must still be tested. Future work will look at much larger networks as well as generalizing the hierarchical definition of clusters for analysis in evolving networks.

The network of campaign contributions, though limited, may prove to be a valuable area of research in the future. Fuzzy hierarchical spectral clustering shows promising results for finding interesting communities at multiple levels and additional experiments and analysis may yield new information and insights into the nature of money in politics. Future use of the data will incorporate some of the other inherent structure of the political landscape, such as lobbyist relationships, directly adding party affiliations, or even adding legislative committee information for candidates. This will allow for further testing with fuzzy heterogeneous networks. Additionally, because of the recent court rulings on campaign finance, it may be more important than ever to track and discover communities within politics.

Acknowledgments

We would like to thank the National Institute on Money in State Politics and their staff for making their data, assistance, and time available for this project.

References

Aynaud, T., and Guillaume, J.-L. 2010. Static community detection algorithms for evolving networks. In *WiOpt'10: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, 508–514.

Bandyopadhyay, S. 2005. Automatic determination of the number of fuzzy clusters using simulated annealing with variable representation. In Hacid, M.-S.; Murray, N.; Ras, Z.; and Tsumoto, S., eds., *Foundations of Intelligent Systems*, volume 3488 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 594–602.

Blondel, V.; Guillaume, J.; Lambiotte, R.; and Mech, E. 2008. Fast unfolding of communities in large networks. *J. Stat. Mech* P10008.

Bonica, A. 2014. Mapping the ideological marketplace. *American Journal of Political Science* 58(2):367–386.

Chauhan, S.; Girvan, M.; and Ott, E. 2009. Spectral properties of networks with community structure. *Phys. Rev. E* 80:056114.

Devillez, A.; Billaudel, P.; and Lecolier, G. V. 2002. A fuzzy hybrid hierarchical clustering method with a new criterion able to find the optimal partition. *Fuzzy Sets and Systems* 128(3):323 – 338.

Flake, G. W.; Lawrence, S.; Giles, C. L.; and Coetzee, F. M. 2002. Self-organization and identification of web communities. *IEEE Computer* 35:66–71.

Fortunato, S. 2009. Community detection in graphs. *CoRR* abs/0906.0612.

Havens, T.; Bezdek, J.; Leckie, C.; Ramamohanarao, K.; and Palaniswami, M. 2013. A soft modularity function for detecting fuzzy communities in social networks. *Fuzzy Systems, IEEE Transactions on* 21(6):1170–1175.

Hopcroft, J.; Khan, O.; Kulis, B.; and Selman, B. 2004. Tracking evolving communities in large linked networks. *Proceedings of the National Academy of Sciences* 101:5249–5253.

Jung, S., and Segev, A. 2014. Analyzing future communities in growing citation networks. *Knowledge Based Systems* 69:34–44.

Liu, J. 2010. Fuzzy modularity and fuzzy community structure in networks. *The European Physical Journal B* 77(4):547–557.

Lusseau, D., and Newman, M. E. 2004. Identifying the role that animals play in their social networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271:S477–S481.

MacCluer, C. 2000. The many proofs and applications of perron's theorem. *SIAM Review* 42(3):487–498.

Newman, M. E. J., and Girvan, M. 2004. Finding and evaluating community structure in networks. *Physical Review* E 69(026113).

Newman, M. 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*.

Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2001. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, 849–856. MIT Press.

Palla, G.; Derényi, I.; Farkas, I.; and Vicsek, T. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435(7043):814–8.

Pons, P., and Latapy, M. 2004. Computing communities in large networks using random walks. *J. of Graph Alg. and App. bf* 10:284–293.

Pothen, A.; Simon, H.; and Liou, K. 1990. Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications* 11(3):430–452.

Power, J. D.; Cohen, A. L.; Nelson, S. M.; Wig, G. S.; Barnes, K. A.; Church, J. A.; Vogel, A. C.; Laumann, T. O.; Miezin, F. M.; Schlaggar, B. L.; and Petersen, S. E. 2011. Functional network organization of the human brain. *Neuron* 72(4):665 – 678.

Sarkar, S., and Dong, A. 2011. Community detection in graphs using singular value decomposition. *Phys. Rev. E* 83:046114.

Sarkar, S.; Henderson, J. A.; and Robinson, P. A. 2013. Spectral characterization of hierarchical network modularity and limits of modularity detection. *PLoS ONE* 8(1):e54383.

Spiliopoulou, M.; Ntoutsi, I.; Theodoridis, Y.; and Schult, R. 2006. Monic: Modeling and monitoring cluster transitions. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, 706–711. New York, NY, USA: ACM.

Torra, V. 2005. Fuzzy c-means for fuzzy hierarchical clustering. In *Fuzzy Systems, 2005. FUZZ '05. The 14th IEEE International Conference on*, 646–651.

Xie, J.; Szymanski, B.; and Liu, X. 2011. Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, 344–349.

Zachary, W. 1977. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* 33:452–473.

Zhang, B., and Horvath, S. 2005. A general framework for weighted gene Co-Expression network analysis. *Statistical Applications in Genetics and Molecular Biology* 4(1):Article 17+.

Zhang, S.; Wang, R.-S.; and Zhang, X.-S. 2007. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. *Physica A: Statistical Mechanics and its Applications* 374(1):483–490.