Community Detection With Katz and Eigenvector Centrality

Complex Networks 2019

Mark Ditsworth[1], Justin Ruths[2]

University of Texas at Dallas

December 10, 2019

[1]markditsworth@protonmail.com
[2]jruths@utdallas.edu

Community Detection

Goal: Given a network, determine which nodes belong to groups based on their network activity

Two Primary Approaches:
Louvain – Iteratively check permutations of community assignments using a modularity maximization heuristic[BGLL08]
Spectral – Find the leading eigenvector of the modularity matrix and transform to discrete community identifiers[New06]

Both approaches can be time and/or resource consuming for large networks.

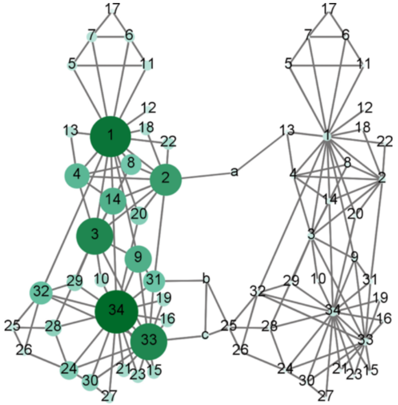
We show another method based on comparing Katz centrality and Eigenvector centrality that is more scalable on large networks.

Localization of Eigenvector Centrality in Modular Networks

Eigenvector centrality ($\mathbf{x}$) calculated by leading eigenvector of adjacency matrix ($\mathbf{A}$)[BL01]

$$\mathbf{x} = \frac{1}{\lambda_1}\mathbf{A}\mathbf{x}$$

Reduced information leads to "localization" of centrality.



Double-karate network with 3-node cut set.
Node size and darkness increase with eigenvector centrality[Sha19]

Robust Katz Centrality Complements Eigenvector Centrality

The following, illustrated in [Sha19], shows that Katz Centrality spans the entire eigenbasis of the adjacency matrix $\mathbf{A}$ of an undirected network.
Katz Centrality[Kat53] is given by,

$$\mathbf{x} = (I - \alpha \mathbf{A})^{-1} \ \beta \mathbf{1} \ , \ \alpha < \frac{1}{\lambda_1}$$

The inverse operation can be expressed as a power series,

$$(I - \alpha \mathbf{A})^{-1} = I + \alpha \mathbf{A} + \alpha^2 \mathbf{A}^2 + \alpha^3 \mathbf{A}^3 + \cdots$$

The vector $\beta \mathbf{1}$ can be expressed as a linear combination of the orthogonal eignvectors of $\mathbf{A}$: $\mathbf{u}_1 \dots \mathbf{u}_n$

$$\mathbf{x} = (I + \alpha \mathbf{A} + \alpha^2 \mathbf{A}^2 + \cdots)(a_1 \mathbf{u}_1 + a_2 \mathbf{u_2} + \cdots + a_n \mathbf{u}_n)$$

Thus, $\mathbf{x} = a_1 \mathbf{u}_1 \sum_{k=0}^{\infty} (\alpha \lambda_1)^k + \cdots + a_n \mathbf{u}_n \sum_{k=0}^{\infty} (\alpha \lambda_n)^k$

Katz-Eigenvector Centrality Plot Identifies Modular Communities

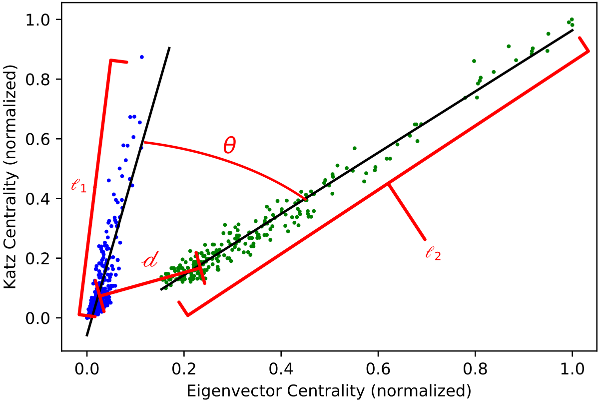
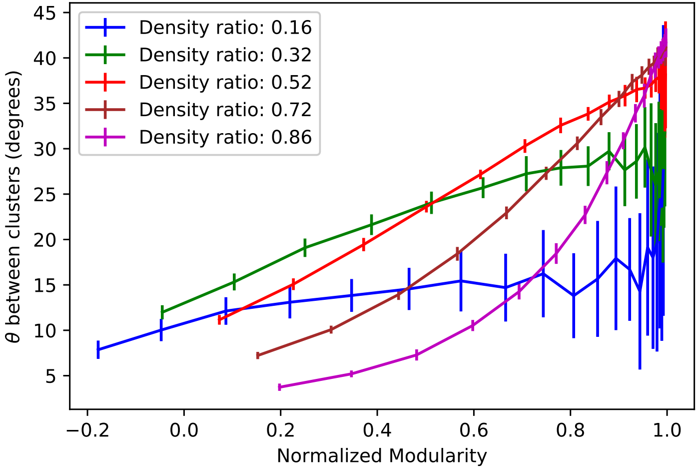
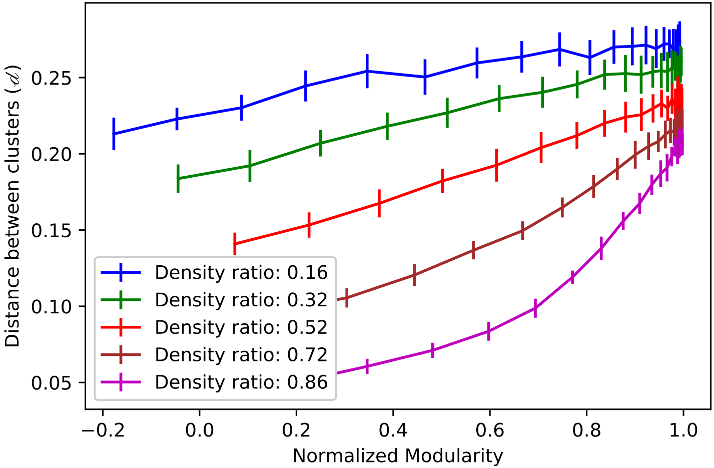


Figure: Ad-hoc modular network with two Barabasi-Albert[BA99] random networks (n=250) joined by 800 randomly placed edges.

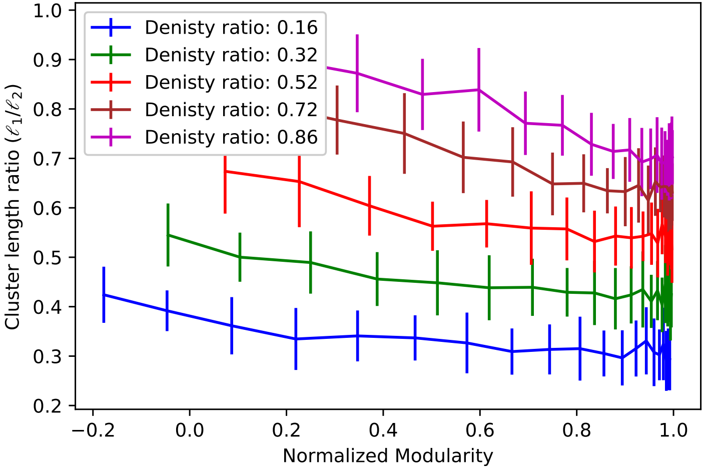
Modularity, Density Affect The KE Plot Clusters' Angle



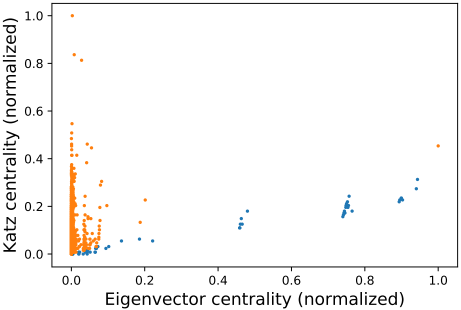
Modularity, Density Affect The KE Plot Clusters' Distance



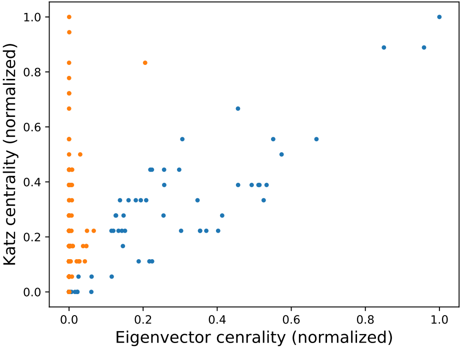
Modularity, Density Affect The KE Plot Clusters' Length Ratio



KE Plot Identifies Communities in Real-World Networks

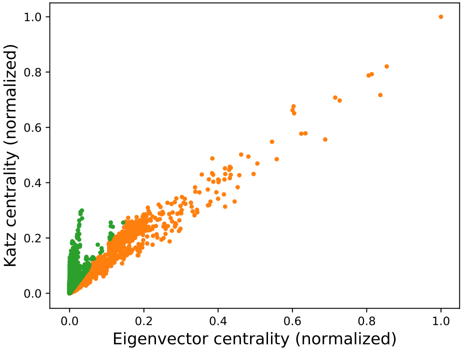


(a) DBLP network[YL12]

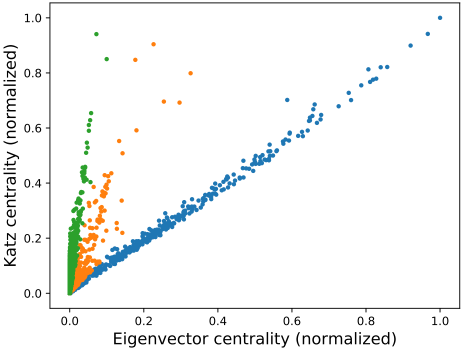


(b) Amazon product network[MTSvdH15]

KE Plot Identifies Communities in Real-World Networks (cont.)



(c) Amazon health network[MTSvdH15]



(d) Amazon beauty network[MTSvdH15]

KE Method Performs Competitively Against Louvain and Spectral

	Runtime		
Network	**L**	**S**	**KE**
AMZN Product	371 ms	231 ms	32 ms
AdHoc BA 1	11.8 s	877 ms	329 ms
AdHoc BA 2	2.03 m	3.88 s	228 ms
DBLP	12.0 m	1.15 hr	751 ms
AdHoc BA 3	2.07 hr	4.65 m	1.97 s
AMZN Beauty	11.7 hr	14.5 m	11.8 s
AMZN Health	16.2 d	5.30 m	35.7 s

Table: Comparison of runtime between Louvain community detection (**L**), spectral community detection (**S**), and the KE plot method of extracting communities from various networks with n nodes and m edges.

KE Method Performs Competitively Against Louvain and Spectral

	Modularity (Q/Q_{max})		
Network	**L**	**S**	**KE**
AMZN Product	0.801/0.908	0.781/0.893	0.359/0.467
AdHoc BA 1	0.485/0.491	0.485/0.490	0.480/0.498
AdHoc BA 2	0.291/0.930	0.454/0.464	0.228/0.471
DBLP	0.805/0.982	0.713/0.974	0.019/0.034
AdHoc BA 3	0.203/0.931	0.382/0.393	0.123/0.492
AMZN Beauty	0.499/0.840	0.566/0.735	0.365/0.645
AMZN Health	0.413/0.543	0.00/0.00	0.423/0.608

Table: Comparison of resulting modularity between Louvain community detection (**L**), spectral community detection (**S**), and the KE plot method of extracting communities from various networks with n nodes and m edges.

KE Method Performs Competitively Against Louvain and Spectral

	Communities		
Network	**L**	**S**	**KE**
AMZN Product	14	17	3
AdHoc BA 1	2	2	3
AdHoc BA 2	17	2	2
DBLP	129	191	2
AdHoc BA 3	18	3	2
AMZN Beauty	25	4	3
AMZN Health	34	1	2

Table: Comparison of the number of detected communities between Louvain community detection (**L**), spectral community detection (**S**), and the KE plot method of extracting communities from various networks with n nodes and m edges.

KE Method Supported By Louvain Results

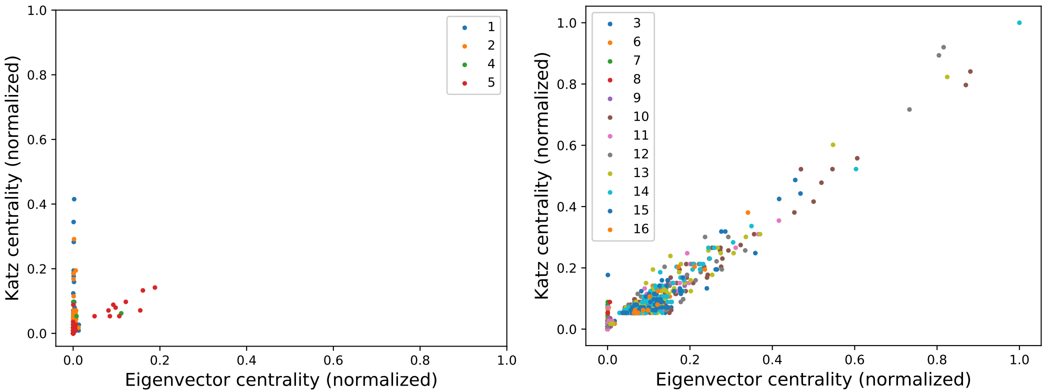
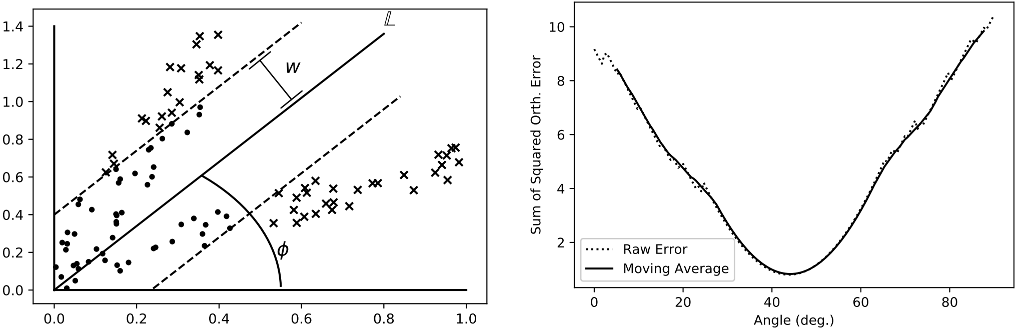


Figure: 16 communities within an ad-hoc modular network detected using Louvain, reduced to two groups that largely follow the pattern utilized by the KE method.

Thank You

Supporting code available at github.com/markditsworth/ModularityStudy

Appendix A: Clustering Algorithm



(a) (b)

Appendix B: KE Plot Example using Erdos-Renyi model

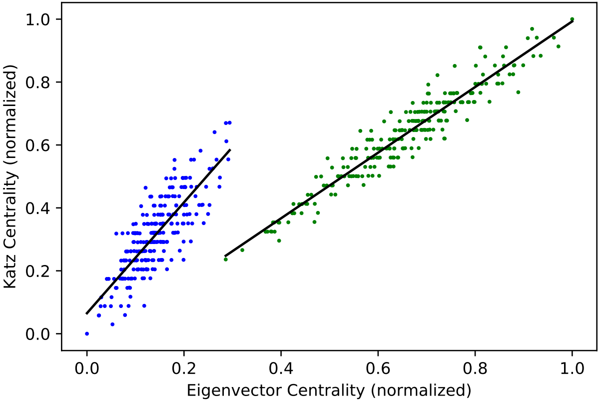
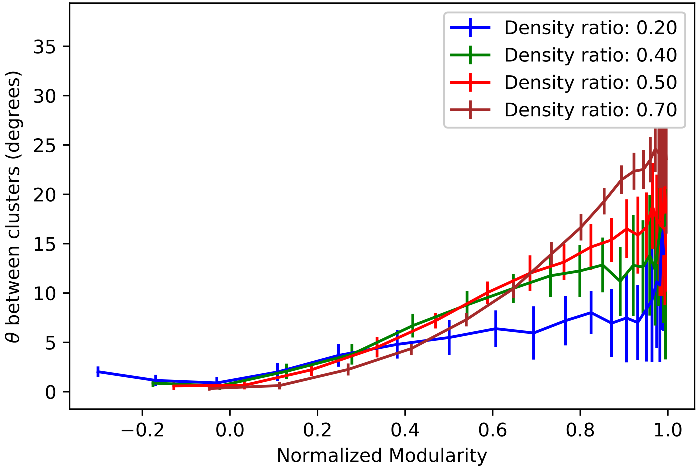
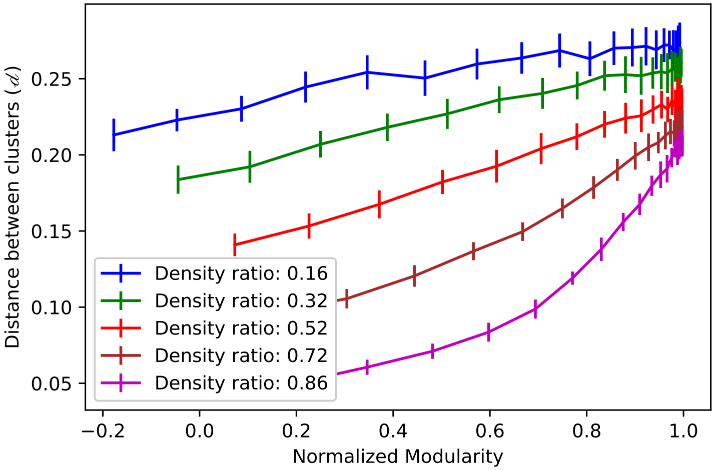


Figure: Ad-hoc modular network with two Erdos-Renyi[ER60] random networks (n=250) joined by 800 randomly placed edges.

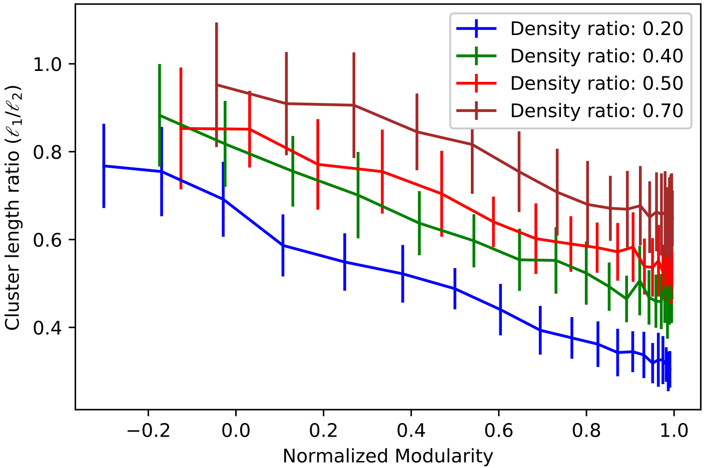
Appendix C: Modularity, Density Affect The KE Plot Clusters' Angle (ER)



Appendix D: Modularity, Density Affect The KE Plot Clusters' Distance (ER)



Appendix E: Modularity, Density Affect The KE Plot Clusters' Length Ratio (ER)



References I

- Albert Lorenz Barabási and R. Albert, *Emergence of scaling in random networks*, Science **286** (1999).
- 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).
- Phillip Bonacich and Paulette Lloyd, *Eigenvector-like measures of centrality for asymmetric relations*, Social Networks (2001).
- Paul Erdős and Alfred Rényi, *On the evolution of random graphs*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5** (1960).
- L. Katz, *A new status index derived from sociometric analysis*, Psychometrika **18** (1953).
- J. McAuley, C. Targett, J. Shi, and A. van den Hengle, *Image-based recommendations on styles and substitutes*, SIGIR, 2015.

References II

- M. E. J. Newman, *Finding community structure in networks using the eigenvectors of matrices*, Phys. Rev. E **74** (2006).
- Kieran J. Sharkey, *Localization of eigenvector centrality in networks with a cut vertex*, Physical Review E **99** (2019).
- J. Yang and J. Leskovec, *Defining and evaluating network communities based on ground-truth*, ICDM, 2012.