CINNA: An R package for deciphering Central Informative Nodes in Network Analysis

Minoo Ashtiani[1], Mehdi Mirzaie[2] and Mohieddin Jafari[1*]

1 Drug Design and Bioinformatics Unit, Medical Biotechnology Department, Biotechnology Research Center, Pasteur Institute of Iran, 69 Pasteur St, PO Box 13164, Tehran, Iran, 2 Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Jalal Ale Ahmad Highway, PO Box 14115-134, Tehran, Iran

*Correspondence:

m_jafari@pasteur.ir

jafareem@gmail.com

jafarilab-pasteur.com

Abstract

Nowadays, reconstructed networks originated from various contexts become more complex and larger, which make them more difficult to figure out. Recognizing influential nodes helps us to comprehend these huge networks in a convenient way. To identify these nodes, several centrality measures based on the network properties are proposed. However, excessive variation of centrality measures complicates the process of choosing appropriate centrality measure for a given network. Therefore, a simple pipeline for comparing these measures and distinguishing which one rightfully points at the central nodes is required.

The CINNA R package conveniently has brought together all required methods for net-work centrality analysis. It contains network component segregation, calculation and prioritizing centralities, along with clustering and visualization functions.

CINNA package is freely available from the R project at http://cran.r-project.org/, http://jafarilab-pasteur.com/content/software/CINNA.html.

1 Background

Networks in the context of different sciences are very enormous, complex and informative. Influence of a node in a network is a consequence of its position which described mathematically using centrality measures. The ultimate goal of centrality analysis is to determine the most influential and basic nodes. Several centrality measures have been introduced to rank nodes based on their importance and detect most influential nodes using different approaches (*1*). Most of these measures can be characterized based on network features including direction and weight. However, the essential concern is, which centrality measures can precisely point at essential nodes of networks.

Many studies have been conducted to compare the behavior of centralities within several networks. Dwyer *et al.* assessed the difference among centrality measures using visualization methods (*2*). At the same time, Borgatti and Everett found that walk structure of a network can strongly affect the centrality measure values (*3*). Thereafter, the correlation among the symmetric and directed issues of common four measures was inquired by Valente *et al.* . They compared 58 various social networks, came to this end that methods for collecting network data sets revolve the correlation across the measures (*4*). At 2014, Batool and Niazi analyzed three different types of neural networks which resulted the close correlation among Closeness-Eccentricity and Degree (*5*). Furthermore, Cong Li *et al.* tried to understand the dependency of several centrality measures based on correlation analysis (*6*). In companion work, we analyzed 27 centrality measures in 14 distinct networks showed that properly detecting central nodes requires utilizing dimensional reduction methods such as principal components analysis (PCA) (*7*). To best of our knowledge, no specified pipeline has been introduced for prioritizing centrality measures within a given net-work. The CINNA R package is a collection of functions which designed to calculate, compare, assort, and visualize network centrality measures together.

2 Functionality

Centrality measure values depend on the topological network features. Therefore, having a general overview of network topology and measurable measures would be a great help to figure out the real central nodes in network analysis (*7*). Most of the centrality measures require strongly connected net-works. Then, functions for segregating components of a network are available for various formats such as igraph (*8*), network (*9*), adjacency matrix or edge list. It is also provided special functions to figure out bipartite networks and apply centrality analysis to one of the projections. In this package, all appropriate measures based on the network structure, are provided including undirected-unweighted, undirected-weighted, directed-unweighted and directed-weighted graph.

In order to distinguish appropriate and most informative centralities, CINNA has prepared PCA (*10*) and t-Distributed Stochastic Neighbor Embedding (t-sne) algorithm (*11*). Both of them are dimensionality reduction approaches respectively in terms of linear and non-linear analysis. In PCA method, using contribution criteria, it can be determined which centrality measures are principal informative and so, which one can describe central nodes of the network more accurately (*7*). On the other hand, t-sne algorithm supplies cost criteria for acquiring most informative centralities. Hence, user can be able to compare computed centrality results in terms of contribution or cost score.

Also, CINNA package provides different visualization techniques for comparing results of centrality analysis and dimension-al reduction methods including heat-map, dendrogram and pair-wise scatter plot. The whole network can be illustrated by specifying a centrality whereas the sizes of vertices indicate the centrality values. The unified manual can be accessed using the R command help (package=CINNA). In addition, an immense step-by-step tutorial of all substantial features of the package can be approached using the command `browseVignette("CINNA")`.

As an example, we used Zachary network (*12*) to briefly glance on CINNA functionality. It can be loaded via `data(zachary)` and is now saved in a variable called `zachary`. The giant component of the network (*13, 14*) can be extracted using `giant.component.extract(zachary)` which is an igraph class (*8*). To declare centrality measures depending on the Zachary network structure, the `proper.centralities` function can be used as follows, `prop.cent=proper.centralities(zachary)`.This function returns a vector including names of centrality measures (44 measures) that are applicable to this network. To compute a set of common centrality measures mentioned in `prop.cent` object, the below function can be used, `calc.cent=calculate.centralities (zachary,except= uncommon.centralities)`. In the next step, using PCA data reduction method an assorted list of measures can be retrieved according to their contribution value to construct principal components as below `pca.centralities(calc.cent)`. The corresponding results are illustrated in Figure 1. As it shown, Closeness centrality (Latora) (*15*) have the highest level of contribution among eight calculated common centrality measures. For accessing the contribution and Eigen values calculated in PCA, the below function can be useful; `summary.pca.centralities(calc.cent)`.

4 Conclusions

The CINNA package currently includes about 23 user-level functions along with 5 natural network examples (*12, 16-19*), which helps the user to have a different experience from the centrality network analysis. An essential part of the advancement process is researchers in network science feedback. We appreciate receiving all suggestions and comments from the users.

Fig. 1. Exemplary uses of CINNA on Zachary network. A) PCA contribution bar plot. B) Heat map of nodes centrality measure values. Colors spectra from blue to red represent nodes that have lowest to highest centrality values. C) Zachary graph visualization. The size of nodes corresponds to the value of Closeness centrality (Latora). D) Association plot between two centrality measures. A Linear relationship between "Closeness centrality (Latora)" and "eigenvector centralities" (*20*) has been shown using a red line. E) Correlation plot among two centrality measures. This plot has four separated parts. Top left and top bottom right belong to respectively "Closeness centrality (Latora)" and "eigenvector centralities". Top right part visualizes the correlation among the variables based on Pearson coefficient. At last, bottom left illustrates scatterplot of variables.

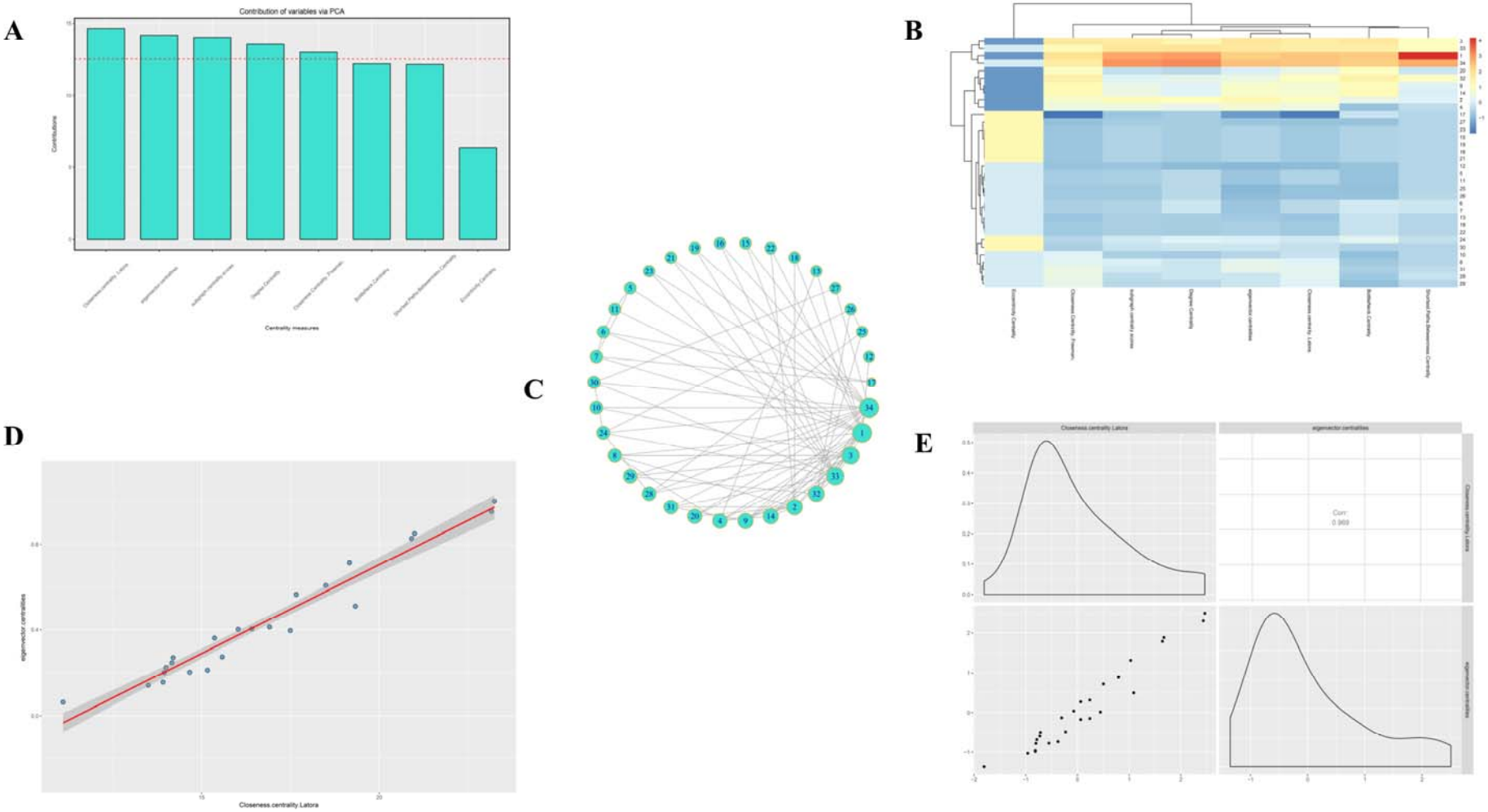
Acknowledgements

The authors appreciate Reza Karbalaei for designing the logo of CINNA package, which can be seen on http://www.jafarilab-pasteur.com/content/software/CINNA.html. Also, we acknowledge Ali Salehza-deh Yazdi and Mahdi Jalili for suggestions and feedback.

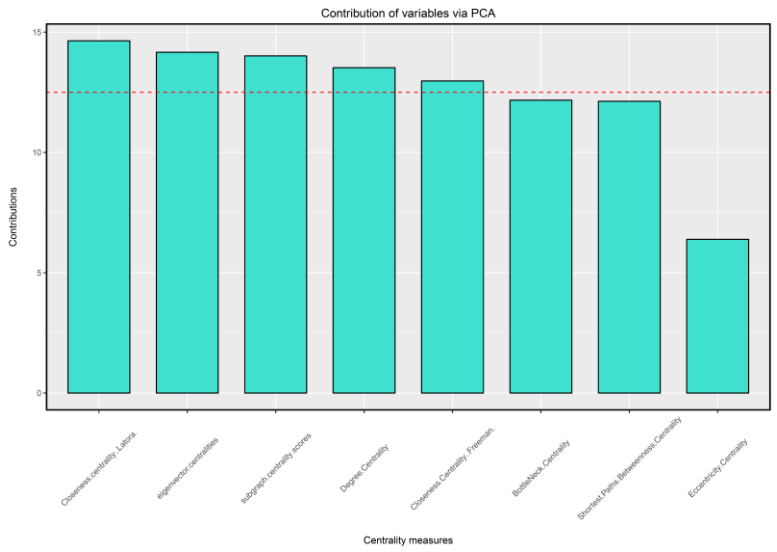
References

1. M. Jalili *et al.*, CentiServer: A Comprehensive Resource, Web-Based Application and R Package for Centrality Analysis. *PloS one* **10**, e0143111 (2015).
2. T. H. Dwyer, S.-H.; Koschützki, D.; Schreiber, F. & Xu, K. , Visual analysis of network centralities Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation. (2006).
3. S. P. Borgatti, M. G. Everett, A graph-theoretic perspective on centrality. *Social networks* **28**, 466-484 (2006).
4. T. W. Valente, K. Coronges, C. Lakon, E. Costenbader, How Correlated Are Network Centrality Measures? *Connect (Tor)* **28**, 16-26 (2008).
5. K. Batool, M. A. Niazi, Towards a methodology for validation of centrality measures in complex networks. *PLoS One* **9**, e90283 (2014).
6. C. Li, Q. Li, P. Van Mieghem, H. E. Stanley, H. Wang, Correlation between centrality metrics and their application to the opinion model. *The European Physical Journal B* **88**, 65 (2015).
7. M. Ashtiani *et al.*, Revisiting the General Concept of Network Centralities: A Propose for Centrality Analysis in Network Science. *bioRxiv*, 149492 (2017).
8. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**, 1-9 (2006).
9. C. T. Butts, network: a Package for Managing Relational Data in R. *Journal of Statistical Software* **24**, 1-36 (2008).
10. H. Abdi, L. J. Williams, Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* **2**, 433-459 (2010).
11. L. v. d. Maaten, G. Hinton, Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**, 2579-2605 (2008).
12. W. W. Zachary, An information flow model for conflict and fission in small groups. *Journal of anthropological research* **33**, 452-473 (1977).
13. M. Newman, *Networks: an introduction*. (Oxford university press, 2010).
14. M. Jafari, M. Sadeghi, M. Mirzaie, S. A. Marashi, M. Rezaei-Tavirani, Evolutionarily conserved motifs and modules in mitochondrial protein-protein interaction networks. *Mitochondrion* **13**, 668-675 (2013).
15. V. Latora, M. Marchiori, Efficient behavior of small-world networks. *Phys Rev Lett* **87**, 198701 (2001).
16. F. Barneh, M. Jafari, M. Mirzaie, Updates on drug-target network; facilitating polypharmacology and data integration by growth of DrugBank database. *Brief Bioinform* **17**, 1070-1080 (2016).
17. D. J. Felleman, D. C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex* **1**, 1-47 (1991).
18. Konect, Kangaroo network dataset. (2016).
19. Konect, Rhesus network dataset. (2016).
20. J. A. Rodríguez, E. Estrada, A. Gutiérrez, Functional centrality in graphs. *Linear and Multilinear Algebra* **55**, 293-302 (2007).

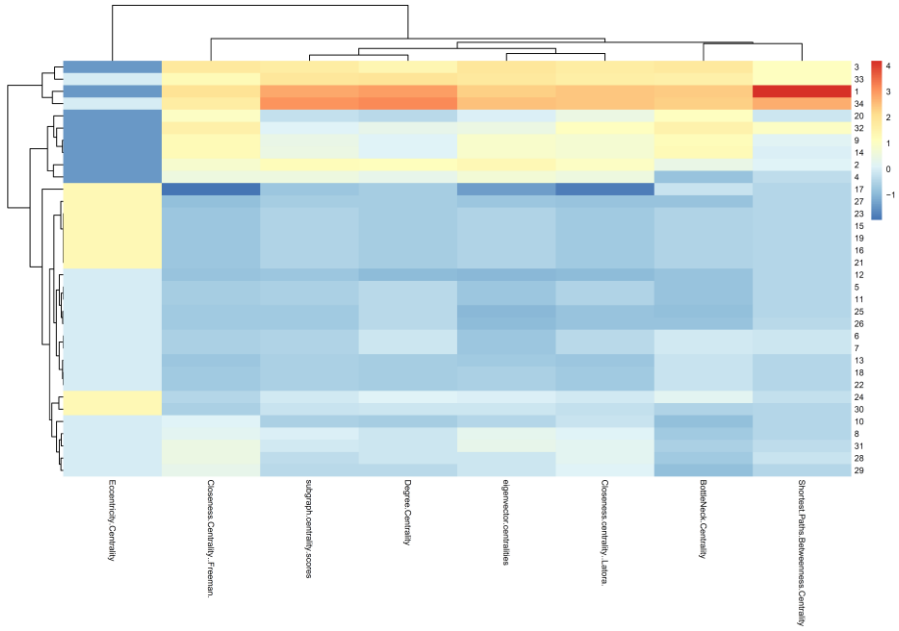
Fig1:


A
Contribution of variables via PCA
Contributions
Centrality measures
B
C
D
eigenvector centralities
Closeness.centrality.Latora
E
Closeness.centrality.Latora
eigenvector centralities
Corr
0.969

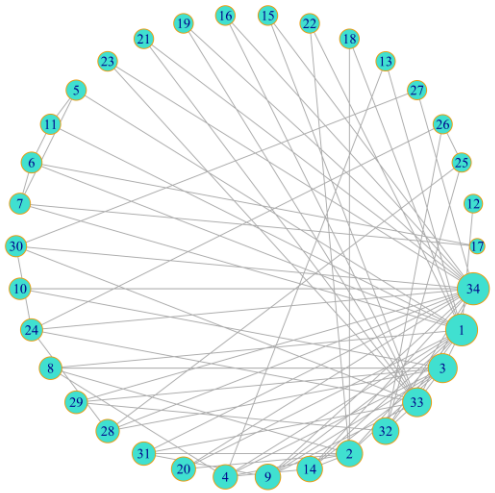
A


Contribution of variables via PCA
Contributions
Centrality measures
Closeness.centrality.Latora
eigenvector.centralities
subgraph.centrality.scores
Degree.Centrality
Closeness.Centrality..Freeman.
BottleNeck.Centrality
Shortest.Paths.Betweenness.Centrality
Eccentricity.Centrality

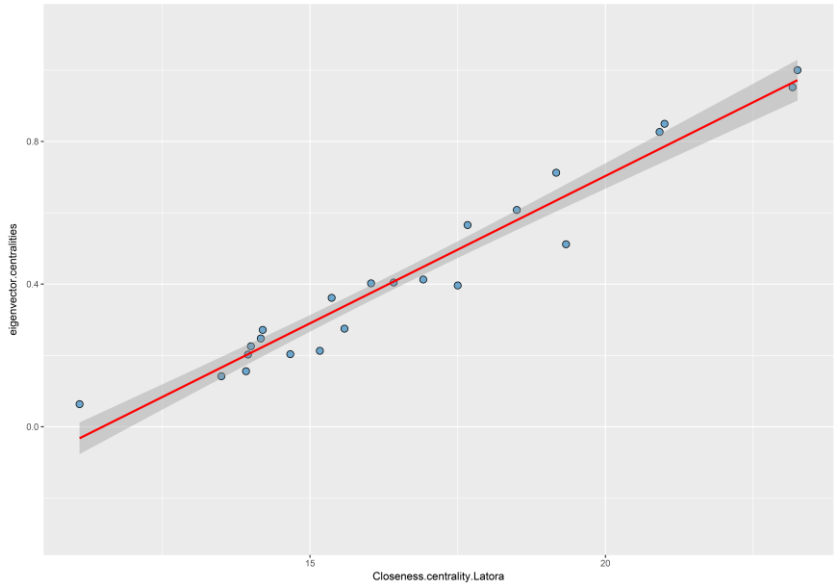
B


Eccentricity.Centrality
Closeness.Centrality..Freeman.
subgraph.centrality.scores
Degree.Centrality
eigenvector.centralities
Closeness.centrality.Latora
BottleNeck.Centrality
Shortest.Paths.Betweenness.Centrality

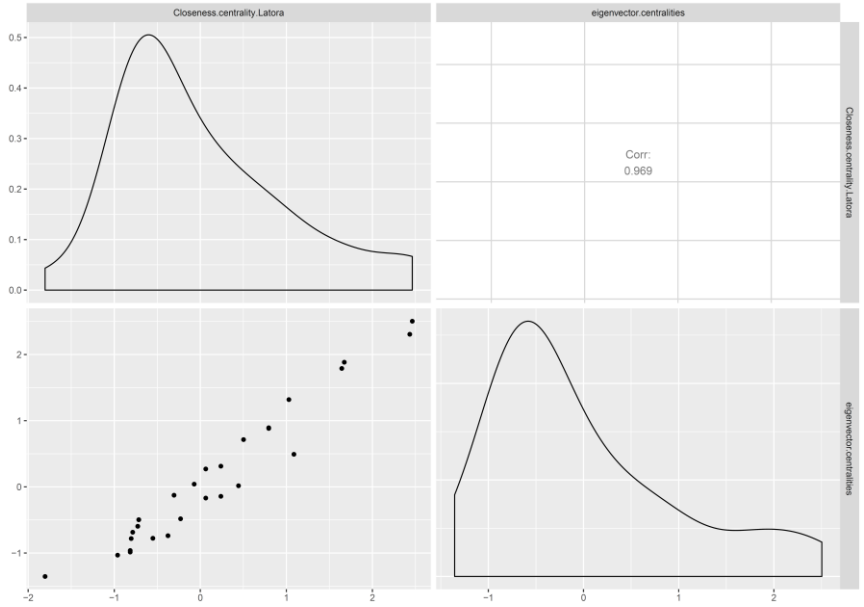
C



D


eigenvector.centralities
Closeness.centrality.Latora

E


Closeness.centrality.Latora
eigenvector.centralities
Corr:
0.969