A guide for choosing community detection algorithms in social network studies: The Question-Alignment Approach

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

Natalie R Smith,[1,2] Paul N Zivich,[2,3] Leah Frerichs,[1] James Moody,[4,5] Allison E Aiello[2,3]

UNC CAROLINA POPULATION CENTER

It is recommended that researchers align their research question and community detection algorithm:
We operationalize that guidance and introduce the ***Question-Alignment Approach***

Community detection is the p**rocess of identifying highly connected subgroups** in a network.

It can be used to examine if health is clustered within communities, overcome analytic challenges, or help target intervention delivery.

Question-Alignment Approach

Step 1:
Clearly define research question

Step 2:
Describe how community detection will be used

Step 3:
Consider the following questions

What biological, social, or behavioral processes are driving the formation of specific communities within a network?

How might the main health variables of interest influence properties of the social network? How are they influenced by the network?

What do ties represent in the network?

Method	Possible use cases
	Divisive
Edge-Betweenness	Identify bridges between communities to disrupt transmission (e.g., of sexually transmitted infections)
	Agglomerative
Walktrap	Transmission between individuals – communication, behaviors, information, or infectious disease
Label Propagation	Adoption of social norms within a network
	Optimization
Infomap	Transmission between individuals – communication, behaviors, information, or infectious disease
Multilevel	Define communities for use as clustering variables in analyses
Spinglass	Define communities for use as clustering variables in analyses

Case study of Question-Alignment Approach:
Step 1: Does an intervention delivered to high-risk communities decrease influenza incidence compared to a broadly disseminated intervention?
Step 2: Use community detection to identify closely connected high-risk groups.
Step 3: We chose Walktrap because of the person-to-person transmission of influenza.

Figure 1: One random individual and their community members

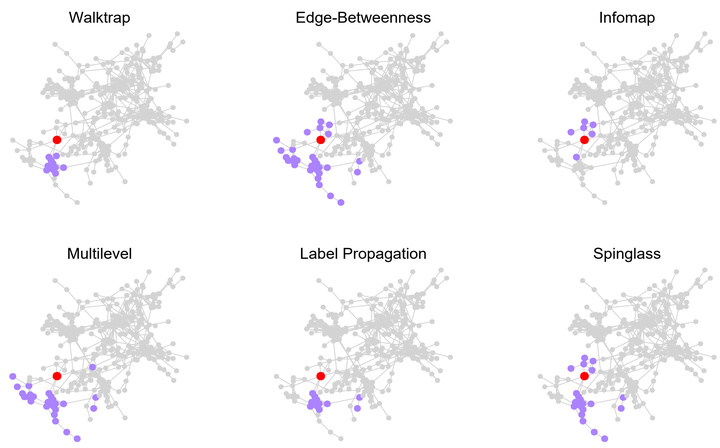


Figure 2: Visual display of communities

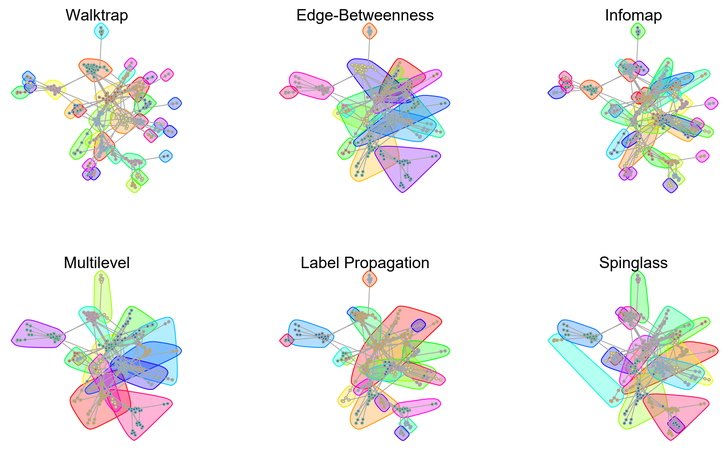
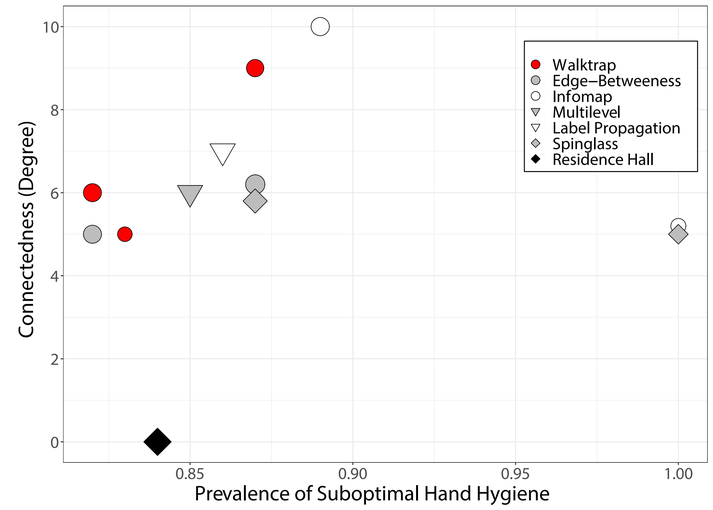


Figure 3: Purposeful selection of communities for intervention



Case study results:

Figures 1 and 2:

Community structure varies greatly between algorithms - total number of communities and individuals within communities

Figure 3:

Walktrap selected the most 'high-risk' communities:

Connected to at least 5 other communities, at least 5 people, and at least 80% of students with suboptimal hand hygiene

Our approach can help researchers obtain the best community structure to answer their research question

Future work should investigate if the Question-Alignment approach improves intervention or analytic results

References:
Yang Z, Algesheimer R, Tessone CJ. A comparative analysis of community detection algorithms on artificial networks. Sci Rep. 2016;6:30750.
Girvan M, Newman ME. Community structure in social and biological networks. Proceedings of the national academy of sciences. 2002;99(12):7821-7826.
Pons P, Latapy M. Computing communities in large networks using random walks. J Graph Algorithms Appl. 2006;10(2):191-218.
Raghavan UN, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. Physical review E. 2007;76(3):036106.
Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences. 2008;105(4):1118-1123.
Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment. 2008;2008(10):P10008.
Reichardt J, Bornholdt S. Statistical mechanics of community detection. Physical Review E. 2006;74(1):016110.

Author Affiliations. 1: Department of Health Policy and Management, UNC Chapel Hill; 2: Carolina Population Center, UNC Chapel Hill; 3: Department of Epidemiology, UNC Chapel Hill; 4: Department of Sociology, Duke University; 5: King Abdulaziz University, Saudi Arabia

Funding. NRS and PNZ received training support from the National Institutes of Health (**T32 HD091058, PI: Aiello**). NRS, PNZ, and AEA received general support from the National Institutes of Health (**P2C HD050924, PI: Frankenberg**). The eX-FLU study was funded by the Centers for Disease Control and Prevention grant **U01CK000185** (AEA). LF was supported by the National Heart, Lung, And Blood Institute of the National Institutes of Health (**K01HL138159**).