HHS Public Access

Author manuscript

Am J Prev Med. Author manuscript; available in PMC 2021 October 01.

Published in final edited form as:

Am J Prev Med. 2020 October ; 59(4): 597–605. doi:10.1016/j.amepre.2020.04.015.

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

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

[1]Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill

[2]Carolina Population Center, University of North Carolina at Chapel Hill

[3]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill

[4]Department of Sociology, Duke University

[5]King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Introduction: Community detection, the process of identifying subgroups of highly connected individuals within a network, is an aspect of social network analysis that is relevant but potentially underutilized in prevention research. Guidance on using community detection methods stresses aligning methods with specific research questions but lacks clear operationalization. The Question-Alignment approach was developed to help address this gap and promote the high-quality use of community detection methods.

Methods: Six community detection methods are discussed: Walktrap, Edge-Betweenness, Infomap, Louvain, Label Propagation, and Spinglass. The Question-Alignment approach is described and demonstrated using real-world data collected in 2013. This hypothetical case study was conducted in 2019 and focused on targeting a hand hygiene intervention to high risk communities to prevent influenza transmission.

Results: Community detection using the Walktrap method best fit the hypothetical case study. The communities derived via Walktrap were quite different from communities derived via the other five methods in both the number of communities and individuals within communities.

There was evidence to support that the Question-Alignment approach can help researchers produce more useful community detection results: compared to other methods of selecting high

Corresponding author information: Natalie R Smith, 1101 McGavran-Greenberg Hall, CB #7411, Chapel Hill, NC, 27599-7411, 412-977-1217, natsmith@live.unc.edu.

Author roles were as follows: NRS conceptualized the approach, ran all analyses, created data visualizations, drafted the initial manuscript, and oversaw all aspects of the work. PNZ provided essential feedback on method development and data visualization throughout the work and drafted detailed mathematical overviews of each algorithm. JM provided expert guidance on the approach and data visualization. LF and AEA provided critical input on writing and interpretation of results. AEA conceptualized, designed, and oversaw the eX-FLU study. All authors provided critical feedback on manuscript drafts and approved the final manuscript.

Financial disclosure: The authors have no financial disclosures.

risk groups, the Walktrap produced the most communities that met the hypothetical intervention requirements.

Conclusions: As prevention research incorporating social networks increases, researchers can use the Question-Alignment approach to produce more theoretically meaningful results and potentially more useful results for practice. Future research should focus on assessing if the Question-Alignment approach translates into improved intervention results.

Introduction

Social networks are related to health.[1-3] Social network analysis has emerged as one tool for evaluating and understanding relationships between individual health and social environments, as well as developing interventions to prevent disease and improve public health.[4-6] A variety of social network analytic methods are available to describe network composition, examine relationships between social networks and health, or identify the underlying group structure of social networks.[2,7]

Identifying groups within social networks is a highly relevant but underutilized tool for public health and prevention research. Typically, due to the diversity of ties in real-world networks, groups within a network are unclear. Community detection, the process of identifying subgroups of highly connected individual within a network, is one popular class of methods. Community detection has been used to examine many public health topics including HIV/AIDS,[8] latrine ownership,[9] smoking cessation,[10] physical activity,[11] cancer treatment patterns,[12] and hospital service regions.[13] Within a social network, communities are groups of individuals who are densely connected to each other, but have few connections to other individuals in the network.[14] Once identified, community structures can help researchers examine if health tends to cluster within communities, or deliver targeted interventions to high-risk communities.[5,14-17] Targeting communities is potentially more appealing than targeting specific individuals because individuals are often more amenable to change when the whole group changes at once, or some interventions may be more naturally targeted towards groups rather than individuals.[5]

There are many community detection methods, or algorithms, available to researchers who are interested in social networks. Algorithms rely on different heuristics to place individuals into mutually exclusive groups. Applying these algorithms is complicated, mainly due to ambiguity around which algorithm is best for a given circumstance.[14,18-22] Recent articles have attempted to provide guidance algorithm selection using criteria such as the mixing parameter of a network, computation time, or overlap with a simulated community structure.[18,23] Other work has noted that the optimal community detection method will depend on how the communities will be used.[22] Indeed, Yang *et al* specifically caution that guidance using criteria like computation time “have to be applied in conjunction with… research questions. A pure application of the recommendations could bias the results.”[18] This appears to be generally acknowledged by community detection researchers,[14,15,18-20,22-24] but concrete demonstrations of how algorithms could be applied in conjunction with health promotion and disease prevention research questions are lacking.

Author Manuscript
Author Manuscript
Author Manuscript

Using a research question to drive algorithm choice allows researchers to leverage each algorithm's different properties to produce a result that best aligns with their question. This paper provides guidance to help prevention researchers with a background in network analysis align their research questions with publicly available community detection algorithms. This method, the Question-Alignment (QA) approach, is intended to help researchers align a research question with the most theoretically appropriate community detection algorithm. This approach may not necessarily lead to better results but is designed to help researchers identify the most theoretically appropriate algorithm for their research question. This method requires a clearly defined research question and an understanding of the specific heuristics employed by algorithms. To help researchers apply the QA approach, the paper provides: (1) an overview of six publicly available algorithms, (2) a discussion of research topics relevant to each algorithm, and (3) an applied example of the QA approach using real-world data.

Algorithms Overview

To begin, overviews for several definitions necessary to understand the described community detection algorithms are provided. *Modularity* quantifies the density of links within communities compared to links between communities (ranges from −1 to 1).[16] *Path lengths* (or 'walks' between two nodes) are the number of edges one would have to use to 'walk' from one node to another. The *shortest path length* is the path between two nodes that involves traveling down the minimal number of edges.[25] A *random walk* is a path between two nodes where each step is randomly chosen.[25] *Dendrograms* are tree diagrams returned by some algorithms that visualize community divisions at each step of the algorithm.[14]

Six popular algorithms are considered. These algorithms are readily available in the R *igraph* package,[26] have shown good computational performance in networks of less than 1000 nodes,[18] and have distinct features: Edge-Betweenness,[27] Random Walktrap,[28] Label Propagation,[29] Infomap,[30] Louvain,[31] and Spinglass.[32] Table 1 provides a high-level overview of each algorithm and outlines features that researchers may be interested in, such as their ability to handle directed networks or be customized. Algorithms are classified as divisive, agglomerative, or optimization based. Further information on each algorithm is available in the Appendix.

Divisive

Divisive algorithms begin with a complete network and iteratively divide the network into smaller communities. The Edge-Betweenness algorithm is one divisive method that defines communities by iteratively removing edges that have a high likelihood of linking separate communities.[27] This algorithm continues until the full network has been completely divided (i.e., each node is its own community) and returns a dendrogram.

Agglomerative

Agglomerative algorithms begin by considering each node as its own community and then iteratively combine nodes into larger communities. The Walktrap method is based on the premise that nodes within communities are likely to be connected by shorter random "walks".[28] Communities are iteratively merged together by minimizing the overall distance between nodes and communities, defined by random walks. The maximum path length of the random walks can be specified to result in more close-knit or more diverse communities. The algorithm continues until all nodes in the network are merged into a single community and returns a dendrogram. Label Propagation identifies communities through the transmission of labels between nodes.[29] Each node starts with a unique label, and randomly selected nodes adopt the label of the majority of its neighbors. The algorithm stops when every node has the same label as the majority of adjacent nodes.

Optimization based

The main feature of optimization algorithms is finding the optimal solution for a pre-specified objective function. The global optimum for modularity can be calculated using cluster_optimal() in the R *igraph* package.[26] Due to the complexity of finding the global optimum for moderately sized networks (100+ nodes), other algorithms, which we focus on here, seek to more efficiently come to a solution. The Louvain algorithm, also known as Multilevel algorithm, is an efficient algorithm that seeks to maximize modularity by merging nodes into communities.[31] The algorithm stops when no merges result in a modularity increase. The Spinglass algorithm optimizes a function that rewards edges with a community and penalizes edges between communities and stops when that function is minimized. That function can be modified to place a bigger (or smaller) emphasis on edges within communities. Finally, the Infomap algorithm focuses on optimizing the flow of information throughout a network.[30] Communities are iteratively merged together to optimize information flow and the algorithm stops when no further optimization is possible.

The Question-Alignment Approach

Similar to prior guidance,[18,19,23] the Question-Alignment approach compels researchers to choose a community detection algorithm based on a specific research question, rather than other criteria such as computation time or convenience. Researchers should first clearly define their research question. Second, they should state how community detection will be used to answer that research question. Then, researchers should answer the following three guiding questions:

- What biological, social, or behavioral process 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 (e.g., physical contact, communication channels, romantic relationships)?

Answers to these questions can then be aligned with and guide selection of a specific community detection algorithm. This approach helps researchers motivate and justify their algorithm selection. When the QA approach suggests multiple algorithms are appropriate, algorithms could be further selected based on stability[33] or researchers could use an ensemble of potential algorithms.[34]

Example Applications of the QA Approach

To help readers apply the QA approach in their own work, specific research topics in the fields of public health and disease prevention that could align with specific algorithms are discussed. Walktrap and Infomap are the two algorithms that define communities based on flow and may be the best suited for questions about transmission of information, communication, behaviors, or infectious disease. Walktrap would be preferred when researchers want flexibility in choosing community structures or the ability to explore several different cut points for communities since Walktrap returns a dendrogram. Infomap may be preferred if a researcher wants to use an algorithm with a clearly defined stopping rule or has a directed network. For example, researchers have attempted to create updated hospital service regions by identifying communities using patient-to-hospital flow data.[13] While these authors used the Louvain algorithm, it may have been more theoretically useful to derive communities using a flow-based algorithm like Walktrap or Infomap.

The Edge-Betweenness algorithm divides the network by iteratively removing edges. The resulting divisions in the dendrogram are different from the dendrogram produced by the Walktrap algorithm due to the different procedures used to identify communities (i.e., divisive vs. agglomerative; edge-betweenness vs. path length). This algorithm is essentially removing 'bridges' between communities, allowing identification of people who link communities. Additionally, Edge-Betweenness indicates how efficiently the network can be split into disconnected pieces. This information is particularly helpful when trying to interrupt transmission within a network. By identifying which edges link groups together, these links can be targeted to prevent wider transmission across groups. The dendrogram output of Edge-Betweenness means these links can be identified at several levels. One potential use would be identifying which connections in a sexual network pose increased risk of transmission across groups underlying the network.

Label Propagation is a useful algorithm for researchers interested in modeling the adoption of social norms or an intervention because the algorithm defines communities based on iterative adoption processes. However, Label Propagation focuses on adoption that occurs when the majority of a specific node's adjacent nodes have adopted a label. Other adoption processes, such as adoption when one adjacent node adopts, or adoption when all adjacent nodes adopt, may not be represented well by this algorithm.

Algorithms that focus on minimizing outside connections while promoting within-community connections, such as Spinglass and Louvain, may be most valuable to researchers interested in using community detection results as part of a larger analytic strategy. After individuals are placed in communities, those communities could be used as clustering variables or fixed effects in standard linear modeling techniques. For example, a recent article identified communities using the Louvain algorithm.[35] The authors then compared the relative contribution of social network communities, schools, and neighborhoods to the total variance in adolescent body mass index in hierarchical linear models.[35] Researchers can also use community detection results to improve their estimation of causal effects. Estimation of causal effects in the presence of interference – when one individual's treatment affects the outcome of their social ties – requires special approaches.

[36] Communities identified via Spinglass or Louvain may provide a good approximation of a structure where interference is present within communities, but not between (i.e., 'partial interference'[36,37]). Community designations can then be used to calculate inverse probability weights extended for settings with interference.[37] In these scenarios, Spinglass may be preferred over Louvain for defining groups best meeting partial interference, since the tuning parameter of Spinglass can place greater emphasis on missing edges between communities, better satisfying the partial interference assumption.

In practice, researchers will also have to consider features of the algorithm and the network under study. For example, certain algorithms are only applicable to undirected networks. Researchers may choose to use an algorithm defined for directed networks or consider symmetrizing their network (i.e., transforming a directed network into an undirected network).

An Example of the QA Approach

The QA approach is demonstrated using an undirected social network from the 2013 eX-FLU social networks and health randomized trial.[38] For clarity, analyses include the largest component of the network, which drops isolated nodes and other small groups on the periphery of the network. This data is used to help answer a hypothetical research question and illustrate how one might apply the QA approach to achieve more relevant analytic results. First, the example research question and algorithm chosen via the QA method are discussed. Then, the results under the chosen algorithm are compared to results under the five other algorithms presented in this paper. All analyses were conducted in 2019.

The illustrative example focuses on preventing influenza transmission with an intervention to promote optimal hand hygiene practices.[39,40] Organizations may be operating under limited resources and cannot afford to broadly disseminate this type of intervention. A researcher might ask: "Does an intervention delivered to high-risk communities decrease the incidence of influenza compared to a broadly disseminated intervention?" This question requires a researcher to identify high-risk groups to target. This can be done using community detection, and the Walktrap algorithm may be the most appropriate algorithm to use because it is most compatible with a person-to-person transmission model.

The Walktrap algorithm does not have a predefined stopping rule, and so there are 314 possible community structures resulting from the algorithm (Figure 1), ranging from all nodes being their own community to all nodes being in one community. The community structure with the highest overall modularity was composed of 35 communities ranging in size from 2 to 44.

Figure 2 compares results obtained using the Walktrap vs. the other five algorithms, using the community division automatically returned by *igraph* for each. Figure 2 highlights the need for purposeful selection of an algorithm because the resulting community structure (i.e., number of communities, distribution of individuals within communities) can differ greatly based on algorithm. Analyses used the adjusted Rand index (AR) to compare each of the five alternative community structures to the Walktrap community structure. The AR index ranges from 0 to 1, with 0 indicating no overlap, and 1 indicating perfect overlap between two structures.[41]-[43] The algorithm that overlapped the most with Walktrap was Infomap (AR=0.72). Edge-betweenness, Spinglass, and Louvain all had similar amounts of overlap with Walktrap (AR=0.68 or 0.69) and Label Propagation had the least overlap with Walktrap (AR=0.55).

Finally, differences in returned community structure may influence the targeted selection of high-risk communities (Figure 3). From the full set of communities identified within each algorithm, communities meeting the high-risk definition can be identified: communities that contain at least five individuals, >80% of whom have suboptimal hand hygiene behaviors (did not wash their hands at least 5 times a day for 20 seconds each[38,44]), and are connected to at least five other groups in the network (i.e., degree centrality[7] above five).

Using Walktrap, there are 3 medium-sized distinctive communities containing 39 people total that are considered high-risk. Those communities may be ideal for intervention delivery, especially because they are also connected to other communities in the network. All other algorithms identify only one or two communities that meet the high-risk criteria. Infomap returns results that are also practically relevant for this case study: one highly connected community with a high prevalence of suboptimal hand hygiene and one smaller community where 100% of individuals have suboptimal hand hygiene. Nonetheless, Walktrap is a better choice because it identifies more communities that meet the high-risk decision criteria and therefore reaches more total individuals (39 via Walktrap vs. 26 via Infomap). The other algorithms identify anywhere between 21 (Louvain) to 42 (Edge-Betweenness) individuals in high-risk communities, but almost all are less connected than the communities identified via the Walktrap algorithm. eX-FLU also collected information on what residence hall the students lived in. As an example of an alternative selection strategy, the residence hall with the highest proportion of suboptimal hand hygiene. There are 40 students in this chosen residence hall, of which 84% have suboptimal hand hygiene practices. This residence hall was considered to have no connections, since it was not selected using social network data. This point in Figure 3 illustrates that intervention groups selected via community detection may be more useful in this theoretical example because they could identify groups with either higher suboptimal hand hygiene prevalence and/or groups that are spread throughout the network.

No single algorithm would be preferred for all research questions. Consider the following alternative questions: “What is the best way to split the network to prevent influenza transmission between groups?” In this case, Edge-Betweenness may be the most useful, since it identifies edges that link communities underlying the network. As another example, consider: “What is the causal effect of optimal hand hygiene on influenza infection?” The Spinglass algorithm might be the most appropriate because it can potentially divide the network into communities more closely corresponding to the assumptions of inverse probability weights for partial interference. As mentioned above, Infomap could also have been used in this scenario. Walktrap was preferred because it returns alternative community divisions that could be examined, and the algorithm can easily be adjusted in R to incorporate a ‘maximum path length’.

Limitations

This approach does not explicitly address how tuning parameters could be incorporated into algorithms. However, adjustments to tuning parameters decrease the ease of implementation and automation of algorithms, which are desirable features for most analyses.[23] One alternative way to customize algorithms results is through dendrograms, which allow researchers to choose community structures other than the structure automatically chosen by R. This paper can provide a useful overview of the algorithms and guidance on their application but should not be considered a sufficient resource to learn the intricacies of community detection algorithms. Finally, the analyses did not demonstrate that the QA approach will necessarily lead to better outcomes for a given study; rather, analyses showed that using the QA approach can help researchers obtain the best community structure to answer their research question.

Conclusions

Community detection is an underutilized method that could have substantial benefits for public health research and developing disease prevention interventions in a network context. Previous work has suggested that researchers align their research question and community detection algorithm choice. This work fills an important gap by operationalizing this suggestion in the form of the Question-Alignment approach. Future research should further refine and specify the QA approach, and use the QA approach to help select community detection algorithms when possible.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The research presented in this paper is that of the authors and does not reflect official policy of the NIH.

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).

References

1. Smith KP, Christakis NA. Social networks and health. Annu Rev Sociol. 2008;34:405–429.
2. Valente TW. Social networks and health: Models, methods, and applications. Oxford University Press; 2010.
3. Heaney CA, Israel BA. Social networks and social support In: Glanz K, Rimer BK, Viswanath K, eds. Health Behavior and Health Education: Theory, research, and practice. Vol 4 San Francisco, CA, US: Jossey-Bass; 2008:189–210.
4. Hunter RF, de la Haye K, Badham J, Valente T, Clarke M, Kee F. Social network interventions for health behaviour change: a systematic review. The Lancet. 2017;390:S47.
5. Valente TW. Network interventions. Science. 2012;337(6090):49–53. [PubMed: 22767921]
6. Valente TW, Palinkas LA, Czaja S, Chu K-H, Brown CH. Social Network Analysis for Program Implementation. PLoS One. 2015;10(6):e0131712. [PubMed: 26110842]

7. Hawe P, Webster C, Shiell A. A glossary of terms for navigating the field of social network analysis. Journal of Epidemiology & Community Health. 2004;58(12):971–975. [PubMed: 15547054]
8. Clémençon S, De Arazoza H, Rossi F, Tran VC. A statistical network analysis of the HIV/AIDS epidemics in Cuba. Social Network Analysis and Mining. 2015;5(1):58.
9. Shakya HB, Christakis NA, Fowler JH. Association between social network communities and health behavior: an observational sociocentric network study of latrine ownership in rural India. Am J Public Health. 2014;104(5):930–937. [PubMed: 24625175]
10. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. Am J Public Health. 2010;100(7):1282–1289. [PubMed: 20466971]
11. Barnett NP, Ott MQ, Rogers ML, Loxley M, Linkletter C, Clark MA. Peer associations for substance use and exercise in a college student social network. Health Psychol. 2014;33(10):1134. [PubMed: 24364375]
12. Pollack CE, Weissman G, Bekelman J, Liao K, Armstrong K. Physician social networks and variation in prostate cancer treatment in three cities. Health Serv Res. 2012;47(1pt2):380–403. [PubMed: 22092259]
13. Hu Y, Wang F, Xierali IM. Automated Delineation of Hospital Service Areas and Hospital Referral Regions by Modularity Optimization. Health Serv Res. 2018;53(1):236–255. [PubMed: 27861822]
14. Fortunato S, Hric D. Community detection in networks: A user guide. Phys Rep. 2016;659:1–44.
15. Fortunato S Community detection in graphs. Phys Rep. 2010;486(3):75–174.
16. Newman ME. Modularity and community structure in networks. Proc Natl Acad Sci U S A. 2006;103(23):8577–8582. [PubMed: 16723398]
17. Shelton RC, Lee M, Brotzman LE, et al. Use of social network analysis in the development, dissemination, implementation, and sustainability of health behavior interventions for adults: A systematic review. Soc Sci Med. 2019;220:81–101. [PubMed: 30412922]
18. Yang Z, Algesheimer R, Tessone CJ. A comparative analysis of community detection algorithms on artificial networks. Sci Rep. 2016;6:30750. [PubMed: 27476470]
19. Sumner KM, McCabe CM, Nunn CL. Network size, structure, and pathogen transmission: a simulation study comparing different community detection algorithms. Behaviour. 2018;155(7-9):639–670.
20. Lancichinetti A, Fortunato S. Community detection algorithms: A comparative analysis. Physical Review E. 2009;80(5):056117.
21. Orman GK, Labatut V. A Comparison of Community Detection Algorithms on Artificial Networks Paper presented at: Discovery Science; 2009//, 2009; Berlin, Heidelberg.
22. Peel L, Larremore DB, Clauset A. The ground truth about metadata and community detection in networks. Science Advances. 2017;3(5):e1602548. [PubMed: 28508065]
23. Dao V-L, Bothorel C, Lenca P. Community structure: A comparative evaluation of community detection methods. 2018.
24. Porter MA, Onnela J-P, Mucha PJ. Communities in networks. Notices of the AMS. 2009;56(9):1082–1097.
25. Newman MEJ. Networks : an introduction. Oxford : Oxford University Press, 2010.; 2010.
26. Csardi G, Nepusz T. R igraph manual pages. 2015; http://igraph.org/r/doc/, 2018.
27. Girvan M, Newman ME. Community structure in social and biological networks. Proceedings of the national academy of sciences. 2002;99(12):7821–7826.
28. Pons P, Latapy M. Computing communities in large networks using random walks. J Graph Algorithms Appl. 2006;10(2):191–218.
29. 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.
30. 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.
31. 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.
32. Reichardt J, Bornholdt S. Statistical mechanics of community detection. Physical Review E. 2006;74(1):016110.

33. Gates KM, Fisher ZF, Arizmendi C, Henry TR, Duffy KA, Mucha PJ. Assessing the robustness of cluster solutions obtained from sparse count matrices. Psychol Methods. 2019;24(6):675–689. [PubMed: 30742473]
34. Dahlin J, Svenson P. Ensemble approaches for improving community detection methods. arXiv preprint arXiv:13090242 2013.
35. Evans CR, Onnela JP, Williams DR, Subramanian SV. Multiple contexts and adolescent body mass index: Schools, neighborhoods, and social networks. Social science & medicine (1982). 2016;162:21–31. [PubMed: 27322912]
36. Hudgens MG, Halloran ME. Toward Causal Inference With Interference. J Am Stat Assoc. 2008;103(482):832–842. [PubMed: 19081744]
37. Perez-Heydrich C, Hudgens MG, Halloran ME, Clemens JD, Ali M, Emch ME. Assessing effects of cholera vaccination in the presence of interference. Biometrics. 2014;70(3):731–744. [PubMed: 24845800]
38. Aiello AE, Simanek AM, Eisenberg MC, et al. Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial. Epidemics. 2016;15:38–55. [PubMed: 27266848]
39. Gould DJ, Moralejo D, Drey N, Chudleigh JH, Taljaard M. Interventions to improve hand hygiene compliance in patient care. Cochrane Database Syst Rev. 2017;9:Cd005186. [PubMed: 28862335]
40. Luangasanatip N, Hongsuwan M, Limmathurotsakul D, et al. Comparative efficacy of interventions to promote hand hygiene in hospital: systematic review and network meta-analysis. BMJ : British Medical Journal. 2015;351:h3728. [PubMed: 26220070]
41. Hubert L, Arabie P. Comparing partitions. Journal of Classification. 1985;2(1):193–218.
42. Moody J Peer influence groups: identifying dense clusters in large networks. Social Networks. 2001;23(4):261–283.
43. Morey LC, Agresti A. The Measurement of Classification Agreement: An Adjustment to the Rand Statistic for Chance Agreement. Educ Psychol Meas. 1984;44(1):33–37.
44. National Center for Emerging and Zoonotic Infectious Diseases. Wash Your Hands. 2019; https://www.cdc.gov/features/handwashing/index.html.
45. Arenas A, Fernandez A, Gomez S. Analysis of the structure of complex networks at different resolution levels. New journal of physics. 2008;10(5):053039.

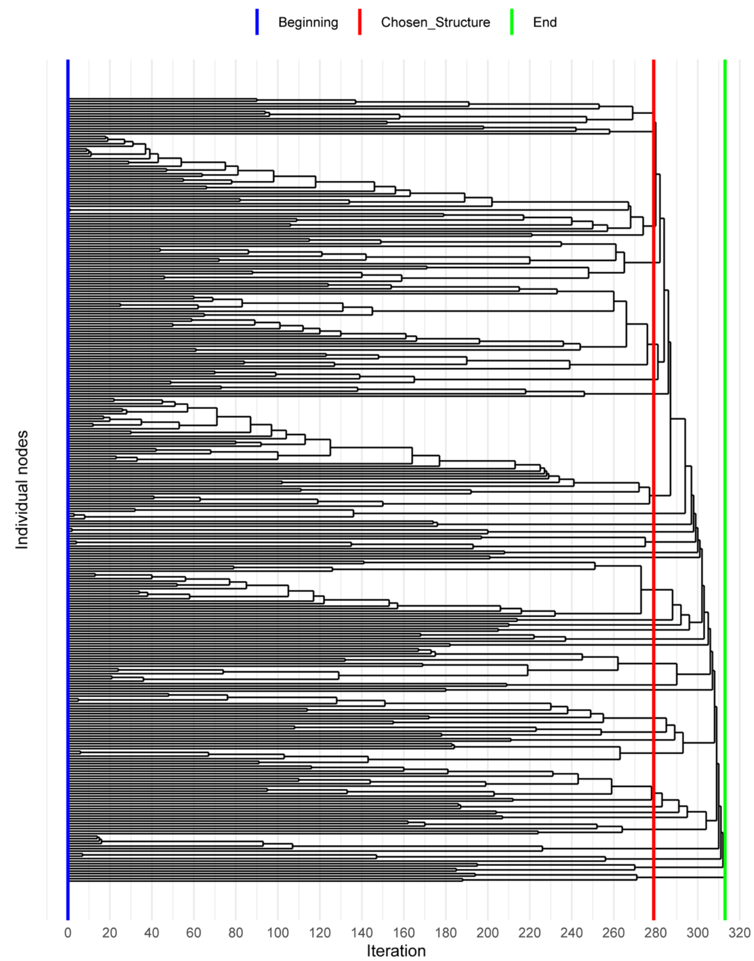

Figure 1:

Dendrogram displaying Walktrap results

Figure notes: Walktrap is an agglomerative algorithm, meaning that each individual node begins as its own community (blue line) and the algorithm ends when all nodes have been merged into one community (green line). Dendrograms are a visual display of the iterative grouping of nodes. In each iteration (x-axis) one node/community is combined with another node/community until all nodes are in one community. Chosen structure is denoted by the red line and is the community division with the highest modularity per *igraph* methods.

Figure 2:

Comparison of community detection results from six algorithms

Figure notes: Colored circles are placed around individuals assigned to the same community. Community membership is mutually exclusive. All results are overlaid on the same network graph.

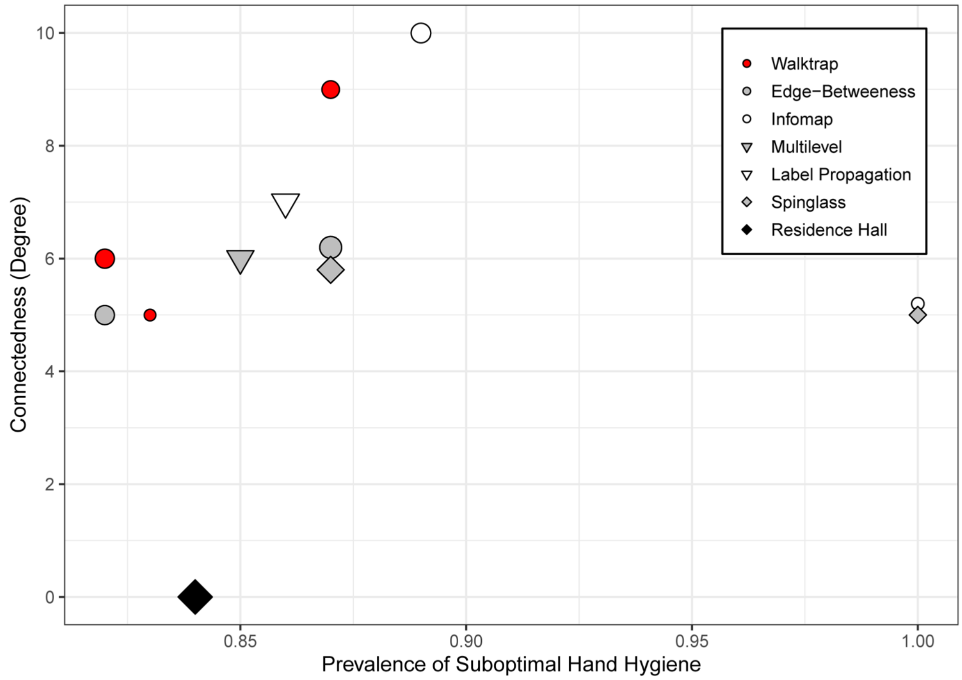

Figure 3:

Communities classified as high-risk by each algorithm

Figure notes: Each community contains at least 5 people, has prevalence of suboptimal hand hygiene that is greater than 80%, and is connected to at least 5 other communities. Selected points are jittered to prevent overlap. Point sizes are scaled by the number of individuals in each community.

Table 1:

Overview of algorithms

Method	Stochastic	Both directed and undirected?	Able to incorporate edge weights[a]	Customization	Possible use cases
Divisive					
Edge-Betweenness[27]		✓	✓	Modify definitions of edges that are likely to link communities	Identify bridges between communities to disrupt transmission
Agglomerative					
Walktrap[28]			✓	Modify allowable length of random walks	Transmission between individuals
Label Propagation[29]	✓		✓		Adoption of social norms
Optimization					
Infomap[30]	✓	✓	✓		Spread of information
Louvain[31]	✓		✓	Add self-loops to increase number of detected communities[b]	Define communities for use as clustering variables in analyses
Spinglass[32]	✓	✓[c]	✓	Tuning parameter (gamma) weights the importance of edges	Define communities for use as clustering variables in analyses

Notes: For all algorithms, igraph will return one community structure. Under algorithms that return dendrograms, is the division with the highest modularity and it is possible to select alternative community structures. In cases of algorithms with specific stopping rules, only one community structure is returned.

[a]If an attribute named ‘weight’ is present in an igraph object, the algorithm will use this by default and the user should supply NA or NULL depending on the algorithm if they do not wish to use it. Researchers should use caution when using edge weights with the Edge-Betweenness algorithm. The algorithm considers weights to be distances, rather than connection strengths (i.e., higher weight = two nodes are farther apart). However, edge weights are considered connection strengths when calculating modularity to determine the final solution (higher weight = stronger connection).

[b]This is not a feature specific to the Louvain algorithm implementation, but rather an alternative method to the Reichardt-Bornholdt resolution parameter that can be used for modularity-based community detection algorithms.[45]

[c]While the algorithm itself allows for directed networks, *igraph* implementation only allows for undirected networks.