# Appendix A:

# Issues in applying common graph-theoretic metrics to stream DAGs

Commonly measured graph properties with potentially limited applicability to stream digraphs can be placed into several classes. These include, but are not limited to, 1) degree-based measures of *centrality* (Borgatti, 2005); 2) graph *distance matrix* summaries; 3) *triangle-based* measures (Luce and Puse 1949); 4) graph spectral summaries, including *eigenvector centrality* (Bonacich 1987), *graph energy* (Gutman & Ramane, 2020), and *hubs and authorities* (Kleinberg 1999); and 5) *tree-width* (Ganian et al., 2016).

## S1-1 Centrality

*Degree centrality* (the nodal indegree and outdegree) will be largely invariant for stream digraphs. All nodes other than sources, sinks, or stream splits or joins will have indegree and outdegree one. This makes nodal metrics based solely on degree, including *coreness* (Seidman 1983) similarly uninformative.

Other common nodal centrality measures require special consideration for describing streams (Table S1). For instance, *betweenness centrality* is calculated as where = number of shortest paths from node to node , = the number of those paths passing through node , and the path term is omitted in the sum if there is no path from node *i* to *j*. Sarker et al. (2019) used this metric for considering the importance of nodes in streams, although streams were represented as undirected graphs. In a stream digraph, betweenness centrality will be highest for stream nodes equidistant from source and sink nodes. In contrast, when considering the *number of reachable nodes from a given node*, source nodes will have the highest values and sink nodes will have the lowest values. This outcome will be reversed when considering the *number of nodes from which a given node is reachable*, with emphasis given to a sink node. Centrality measures can be extended into global measures of *centralization* (Freeman 1978):

where *ci* is the *i*th nodal centrality score of interest (e.g. degree centrality, betweenness centrality, etc.) for a digraph, *D*, and is the maximum centrality outcome for nodes in *D*. We have found that this measure produces global outcomes that tend to increase as a stream DAG becomes increasingly disconnected, at least for *degree centralization*. Therefore, *C*(*D*) may not reflect conventional hydrological conceptions of network connectedness.

## S1-2 Distance matrix summaries

A large number of global graph metrics are computed using the graph distance matrix, ***D***. These include the average distance, the Weiner index (Weiner 1949) the hyper-Wiener index (Randić 1993), *q*-analogs of the Weiner index (Zhang et al. 2012), the Wiener polynomial (Hosoya 1988), and the Tratch–Stankevich–Zeﬁrov index (Tratch et al. 1990). Distance-matrix-based metrics are problematic for stream DAGs because disconnected and upstream nodal distances will be infinitely large. Several methods, including the Harary index (Plavšić et al. 1993) and the Balaban *J* index (Balaban 1982), address this problem by using sums of *reciprocal* distances, also called *efficiencies* (Latora and Marchiori 2001), where the reciprocal distance is taken to be zero. Ivancuic et al. (2000) compare global graph summaries based on ***D***, the reciprocal distance, the distance-path, the reciprocal distance-path, the path Szeged, and the reciprocal path Szeged matrices. Infinite internodal distances are also conceptually problematic for *closeness centrality* (Freeman 1977), the average distance of the shortest paths between a node and all other nodes.

## S1-3 Graph triangles

A number of global graph metrics, including *reciprocity* (Garlaschelli and Loffredo 2004)*,* and the *clustering coefficient*, also called *transitivity* (Holland and Leinhardt 1971), are based on the number of graph *triangles*, i.e., the number of neighbors of a node that are also neighbors of each other. These measures are unlikely to be useful in the surface networks of streams, because of the rarity of triangles in these systems, but may be important in conceptual subsurface stream networks.

## S1-4 Graph spectral summaries

### S1-4.1 Eigenvector centrality

*Eigenvector centrality* (Bonacich 1987), a local (nodal) graph metric, is the principal eigenvector of the graph adjacency matrix. This analysis is more complicated in DAGs than in undirected graphs because the DAG adjacency matrix will be asymmetric, and consequently have distinct left- and right-hand eigenvectors. Other issues include the fact that source nodes, which must have indegree zero, will drive all downstream nodes to have an eigenvector centrality of zero (Newman 2018, pg. 162).

### S1-4.2 Graph energy

*Graph energy*, a global graph metric, is the sum of the absolute values of the eigenvalues of the adjacency matrix, ***A*** (Gutman 1978). From this basic framework more than one hundred graph energy variants have been developed, generally based on different matrix representations of graphs (Gutman and Furtula 2019). High graph energies may indicate unstable graph structures (see Estrada and Benzi 2017). *Graph stability* can also be assessed by evaluating the maximum and minimum eigenvalues of the adjacency matrix with respect to a *master stability function* (Newman 2018). However, because the adjacency matrix of a DAG will be asymmetric, it need not be diagonalizable, and some eigenvalues of the adjacency matrix may be strictly complex. Efforts have been made to extend graph energy to digraphs by using real parts of eigenvalues (Pena and Rada 2008), a skew adjacency matrix and the modulus of the resulting purely imaginary eigenvalues (Adiga et al. 2010), and/or singular values in place of eigenvalues. The *Laplacian matrix*, ***L***, of a digraph (along with ***L*** variants) will also be asymmetric, thus complicating global graph metrics based on eigenanalysis. These include *algebraic connectivity*, i.e., the largest nonzero eigenvalue of ***L*** (Fiedler 1973), and the *Laplacian spectral radius*, i.e., the principal eigenvalue of ***L***.

### **S1-4.3** Hubs and Authorities

The *Hyperlink-Induced Topic Search* algorithm (*HITS*; Kleinberg 1999), also known as *hubs and authorities*, provides a method for ranking the importance of network nodes. Under the *HITS* framework, hubs refer to nodes that point to the most important network nodes, which are in turn the authorities. Hub scores are the entries in the principal eigenvector of where is the graph adjacency matrix of the graph, while authority scores are the entries in the principal eigenvector of . The *HITS* approach has been recommended as a potentially superior alternative to eigenvector centrality (Newman 2018). However, due to asymmetry there may not be a unique direction for the eigenvectors of ***A*** and as a result an eigenanalysis algorithm can return different values when applied to the same network. A potential solution to the problem of asymmetric stream DAG adjacency and Laplacian matrices is to consider the underlying undirected graph. However, this may lead to illogical inferences for streams because flow direction is in fact *not* reversible.

## S1-5 Tree-based measures

In their fully wetted form (and in the absence of splits, resulting in islands) intermittent stream digraphs represent a subclass of DAGs called *directed trees* (Deo 2017, pg. 206) or *oriented trees* (Simion 1991)*.* In a *directed tree*, if oriented arcs are replaced with undirected arcs, the result is a graph that is both connected and acyclic. Additionally, because stream systems flow from sources to the network sink(s), fully wetted intermittent stream DAGs (without islands) can be considered *anti-arborescences* or *directed in-trees* (Deo 2017, pg. 207) a subclass of *rooted tree* graphs, with an orientation *towards* the root (sink).

Beyond simply considering the tree-width of the underlying undirected graph (e.g., Berwanger et al., 2012), there are no reasonable generalizations of tree-width to DAGs that preserve its algorithmically useful and structural properties (Ganian et al., 2016). Furthermore, tree-width-based global measures quantify *departure* from a tree graph structure, but the topology of stream DAGs will be treelike (Table S1). Thus, noninformative tree-width type measures of DAGs include *treewidth* (Robertson & Seymour, 1986) and *branch-width* (Robertson & Seymour, 1991), which require undirected graphs, and *DAG-width* (Berwanger et al., 2012), *Kelly-width* (Hunter & Kreutzer, 2008), and *tree girth* (Bang-Jensen & Gutin, 2007), which will be invariant for acyclic graphs (Ganian et al., 2016).

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Description | Local or Global | Issue(s) |
| Degree centrality | Nodal in and/or outdegree | Local | Invariant. Nodes other than splits and the sink will all have outdegree one. |
| Coreness (Seidman 1983) | The *k*-core of a graph is a maximal subgraph in which each vertex has at least degree *k.* | Global | Invariant, because stream DAG node degrees are largely invariant. |
| Betweenness centrality (Freeman 1977) | (see text for definitions) | Local | Will potentially overemphasize nodes equidistant from source and sink nodes. |
| Centralization (Freeman 1979) | (see text for definition) | Global | Conceptions of centrality may be counter to hydrological conceptions of connectivity |
| Closeness centrality (Freeman 1977) | Average length of the shortest path between a node and all other nodes | Local | Applies only to strongly connected graphs with noninfinite nodal distances. |
| Eigenvector centrality (Bonacich 1987) | See Newman (2018) | Local | Requires a symmetric adjacency matrix. |
| Average nodal distance, Wiener index (Wiener, 1947), hyper-Wiener index (Randić 1993), *q*-analogs of the Wiener index (Zhang et al. 2012), Wiener polynomial (Hosoya 1988),Tratch–Stankevich–Zeﬁrov index (Tratch et al. 1990). | Summaries of the graph distance matrix, ***D*** | Global | Disconnected and upstream distances will be infinitely large. |
| Reciprocity (Garlaschelli and Loffredo 2004) and Transitivity (Holland and Leinhardt 1971) | Occurrence of nodes in triangles | Global | Triangles will be extremely rare in stream DAGs. |
| Graph energy (Gutman, 1978) | Sum of the absolute values of the eigenvalues of the adjacency matrix | Global | Requires a symmetric adjacency matrix |
| Digraph energy (Pena & Rada, 2008), Skew energy of a digraph (Adiga et al., 2010) | See descriptions in this Appendix | Global | Will always equal 0 for DAGs (Pena & Rada, Example 2.4). |
| Algebraic connectivity (Fiedler 1973), Laplacian spectral radius | Eigenvalues from ***A*** or ***L*** | Global | Requires a symmetric adjacency matrix. |
| *HITS* (Kleinberg 1999) | Eigenvectors of | Local | Instability of algorithm. |
| Treewidth (Robertson & Seymour, 1986), Branch-width (Robertson & Seymour, 1991) | Divergence from an acyclic tree | Global | Designed for undirected graphs. |
| DAG-width | See Berwanger et al., (2012) | Global | Will always equal 0 for DAGs (Ganian et al. 2016, following Definition 3.3). |
| Kelly-width | See Hunter & Kreutzer (2008) | Global | Will always equal 1 for DAGs (Ganian et al. 2016, following Definition 3.3). |
| Tree girth | Length of the shortest cycle in a digraph (Bang-Jensen & Gutin 2007, pg 329) | Global | Set equal to for DAGs (Bang-Jensen & Gutin 2007, pg. 12). |

Table S1. Common graph measures that may be poorly suited to stream digraphs or require special consideration.

# Bibliography

Adiga, C., Balakrishnan, R., & So, W. (2010). The skew energy of a digraph. *Linear Algebra and Its Applications*, *432*(7), 1825–1835.

Balaban, A. T. (1982). Highly discriminating distance-based topological index. *Chemical physics letters*, *89*(5), 399-404.

Bang-Jensen, J., & Gutin, G. (2007). Digraghs: Theory, algorithms and applications. *Springer Monographs in Mathematics, Springer-Verlag London Ltd., London*, *101*.

Berwanger, D., Dawar, A., Hunter, P., Kreutzer, S., & Obdržálek, J. (2012). The dag-width of directed graphs. *Journal of Combinatorial Theory, Series B*, *102*(4), 900–923.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, *92*(5), 1170-1182.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71.

Buckley, F., & Harary, F. (1990). *Distance in graphs* (Vol. 2). Redwood City: Addison-Wesley.

Deo, N. (2017). *Graph theory with applications to engineering and computer science*. Courier Dover Publications.

Estrada, E., & Benzi, M. (2017). What is the meaning of the graph energy after all? *Discrete Applied Mathematics*, *230*, 71-77.

Fiedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, *23*(2), 298-305.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41.

Freeman, L. C. (1978). Centrality in social networks, Conceptual clarification. *Social networks*, *1*(3), 215-239.

Garlaschelli, D., & Loffredo, M. I. (2004). Patterns of link reciprocity in directed networks. *Physical review letters*, *93*(26), 268701.

Ganian, R., Hliněnỳ, P., Kneis, J., Meister, D., Obdržálek, J., Rossmanith, P., & Sikdar, S. (2016). Are there any good digraph width measures? *Journal of Combinatorial Theory, Series B*, *116*, 250–286.

Gutman, I. (1978). The energy of a graph, *Ber. Math–Statist. Sekt. Forschungsz. Graz* 103:1-22.

Gutman, I., & Furtula, B. (2019). Graph energies and their applications. *Bulletin (Académie serbe des sciences et des arts. Classe des sciences mathématiques et naturelles. Sciences mathématiques)*, (44), 29-45.

Gutman, I., & Ramane, H. (2020). Research on graph energies in 2019. *MATCH Commun. Math. Comput. Chem*, *84*(2), 277–292.

Holland, P. W., & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative group studies*, *2*(2), 107-124.

Hosoya, H. (1988). On some counting polynomials in chemistry. *Discrete Applied Mathematics*, *19*(1-3), 239-257.

Hunter, P., & Kreutzer, S. (2008). Digraph measures: Kelly decompositions, games, and orderings. *Theoretical Computer Science*, *399*(3), 206–219.

Ivanciuc, O., Ivanciuc, T., Cabrol-Bass, D., & Balaban, A. T. (2000). Evaluation in quantitative structure− property relationship models of structural descriptors derived from information-theory operators. *Journal of Chemical Information and Computer Sciences*, *40*(3), 631-643.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, *46*(5), 604-632.

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, *87*(19), 198701.

Luce, R. D., & Perry, A. D. (1949). A method of matrix analysis of group structure. *Psychometrika*, *14*(2), 95-116.

Newman, M. E. (2018). *Networks, 2nd edition*. Oxford University Press.

Pena, I., & Rada, J. (2008). Energy of digraphs. *Linear and Multilinear Algebra*, *56*(5), 565–579.

Plavšić, D., Nikolić, S., Trinajstić, N., & Mihalić, Z. (1993). On the Harary index for the characterization of chemical graphs. *Journal of Mathematical Chemistry*, *12*(1), 235–250.

Randić, M. (1993). Novel molecular descriptor for structure—property studies. *Chemical Physics Letters*, *211*(4-5), 478-483.

Robertson, N., & Seymour, P. D. (1986). Graph minors. II. Algorithmic aspects of tree-width. *Journal of Algorithms*, *7*(3), 309–322.

Robertson, N., & Seymour, P. D. (1991). Graph minors. X. Obstructions to tree-decomposition. *Journal of Combinatorial Theory, Series B*, *52*(2), 153–190.

Sarker, S., Veremyev, A., Boginski, V., and Singh, A. (2019). Critical nodes in river networks. *Scientific Reports*, *9*(1), 1–11.

Seidman, S. B. (1983). Network structure and minimum degree. *Social networks*, *5*(3), 269-287.

Simion, R. (1991). Trees with 1-factors and oriented trees. *Discrete Mathematics*, *88*(1), 93-104.

Tratch, S. S., Stankevitch, M. I., & Zefirov, N. S. (1990). Combinatorial models and algorithms in chemistry. The expanded Wiener number—a novel topological index. *Journal of computational chemistry*, *11*(8), 899-908.

Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, *69*(1), 17–20.

Zhang, Y., Gutman, I., Liu, J., & Mu, Z. (2012). q-Analog of Wiener index. *Match-Communications in Mathematical and Computer Chemistry*, *67*(2), 347.