Supplementary Information: Sequential detection of temporal communities by estrangement confinement

Vikas Kawadia
vkawadia@bbn.com
Raytheon BBN Technologies, Cambridge MA 02138

Sameet Sreenivasan
sreens@rpi.edu
Social and Cognitive Networks Academic Research Center,
Rensselaer Polytechnic Institute, Troy NY 12180

Interpretation of estrangement confinement as random walk temporal stability

Here we show that estrangement has a principled interpretation in the form of a connection to the notion of *stability* of random walks introduced by Lambiotte et al. [1], and further investigated by Mucha et al. [2]. Stability compares the probability that a random walker at stationarity is in the same community after a step, with the analogous probability obtained for a random walker on a degree-sequence preserving null model of the network under consideration, yielding the following expression:

$$S_{\text{structural}} = \frac{1}{2M} \sum_{uv} \left(A_{uv} - \frac{k_u k_v}{2M} \right) \delta(l_u, l_v)$$

Here, we refer to stability as defined in [1] as *structural* stability as it is a function solely of the structure of the network. $S_{\text{structural}}$ is identical to modularity as shown in [1].

We extend the notion of stability to temporal networks by incorporating an additional term which characterizes the *temporal stability* of a random walk. The temporal stability (of a partition P_t) compares the probability that a random walker in G_t at stationarity, walks along a *historical edge* – an edge in G_t that was an intra-community edge at $t-1$ – and ends up in the same community, with the value of this probability obtained in the maximally temporally stable case viz. the case where the chosen partition for G_t makes every historical edge an intra-community edge. Thus, temporal stability measures the degree to which a random walker's environment remains invariant (i.e. it is in the same community) after a one step walk in t, given that it was invariant for a one step walk in $t-1$.

Formally, the temporal stability as defined above can be written as:

$$S_{\text{temporal}} = \sum_{uv} \frac{Z_{uv}}{2M} \delta(l_u, l_v) - \sum_{uv} \frac{Z_{uv}}{2M}$$

where l_u, l_v are defined by the partition P_t under consideration, and Z_{uv} is as defined in Eq. 3 in Results in the main paper. Here, the second term is the one obtained for the maximally temporally stable case. As is clear from the definition of E (Eq. 3 of main paper), $S_{\text{temporal}} = -E$. Thus, $S_{\text{temporal}} \leq 0$. It follows that the constrained optimization problem to be solved for finding temporal communities – Eq. 1 in the main paper – is equivalent to the problem of maximizing structural stability, $S_{\text{structural}}$, while constraining temporal stability, S_{temporal}, to be greater than or equal to $-\delta$ (where δ is non-negative). Thus, estrangement confinement has a fundamental interpretation in terms of the properties of a random walk on the evolving network.

Estrangement and overlapping communities

Communities in networks often overlap such that nodes can simultaneously belong to multiple groups. Methods for uncovering overlapping communities in static networks have been recently proposed by Ahn et al. [3], and Evans and Lambiotte [4]. Temporal communities can also reveal overlapping communities in the aggregate network comprising all snapshots, since a node can participate in multiple communities over time. The definition of estrangement is easily generalized to the case where even within a snapshot nodes may belong to multiple overlapping communities. If the overlapping community membership of a node in a given snapshot is represented by a set of labels, we first define the *consort score* of an edge as the Jaccard similarity of label sets of the endpoint nodes. Estrangement is then defined as the sum over all edges, of the difference in consort score of an edge from time t to $t-1$, divided by the number of edges. This definition clearly reduces to the definition in the main paper if the label sets are of size one, i. e. , the communities are non-overlapping. Note that this generalized definition of estrangement is decomposable into node-local components as well, and thus – similar to our method for non-overlapping communities – methods for finding overlapping communities can be adapted to find overlapping temporal communities by constraining this generalized estrangement.

The Lagrangian L is unaffected by the induce-graph operation

In HLPA, after the local update rule (Eq. 6 in main paper) has converged, a new graph is *induced* from the current graph and the partition that the update rule has converged to. In this new graph, the communities of the converged partition play the role of the nodes, which we call "supernodes". Each supernode in the induced graph has a self-loop with a weight equal to twice the sum of weights of all links in the original graph contained in the community that forms the supernode. Similarly, a link between two supernodes has a weight equal to the sum of the weights of all edges connecting the two communities represented by the supernodes in the original graph.

Here we prove that the modularity Q and the estrangement E computed for a partition of an induced graph yields the same result as that computed when considering all nodes and edges within the supernodes of the induced graph. As a result, the Lagrangian computed for a partition on the induced graph is also identical to that computed when taking into account all nodes and edges contained within them. We use indices u,v to refer to supernodes of an induced graph, and i,j to refer to nodes of the original graph (contained within supernodes of the induced graph). Thus u and v also refer to community labels at the previous hierarchical level. For a given partition of the induced graph the modularity can be written as :

$$Q^{\text{induced}} = \sum_c \sum_{u,v \in c} \left(\frac{A_{uv}}{2M} - \frac{k_u k_v}{(2M)^2} \right)$$

where c runs over indices of the different communities. We can write A_{uv} as $\sum_{i \in u, j \in v} A_{ij}$. Similarly, $k_u = \sum_{i \in u} k_i$ and $k_v = \sum_{j \in v} k_j$ which gives:

$$Q^{\text{induced}} = \sum_c \sum_{u,v \in c} \left(\sum_{i \in u, j \in v} \frac{A_{ij}}{2M} - \sum_{i \in u, j \in v} \frac{k_i k_j}{(2M)^2} \right)$$

By transitivity of the community labels (i.e. $i \in u$ and $u \in c \implies i \in c$), we can therefore write:

$$Q^{\text{induced}} = \sum_c \sum_{i,j \in c} \left(\frac{A_{ij}}{2M} - \frac{k_i k_j}{(2M)^2} \right)$$

which is identical to the modularity of the same partition computed over nodes of the original graph. Similarly, the term which accounts for the contribution from estrangement due to a partition of the induced graph can be written as:

$$\sum_{u,v} \frac{Z_{uv}}{2M} \delta(l_u, l_v) = \sum_c \sum_{u,v \in c} \frac{Z_{uv}}{2M}$$

where, as before, c runs over the indices of the different communities. We can write $Z_{uv} = \sum_{i \in u, j \in v} Z_{ij}$. Thus the estrangement term becomes:

$$\sum_c \sum_{u,v \in c} \sum_{i \in u, j \in v} \frac{Z_{ij}}{2M} = \sum_c \sum_{i,j \in c} \frac{Z_{ij}}{2M} = \sum_{ij} Z_{ij} \delta(l_i, l_j)$$

which is identical to the estrangement term computed on the original graph. It follows that for a given partition the Lagrangian is preserved when moving from the original graph to the induced graph.

References

[1] Lambiotte, R., J.-C., D. & M., B. Laplacian dynamics and multiscale modular structure in networks. arXiv:0812.1770v3 (2010).

[2] Mucha, P., Richardson, T., Macon, K., Porter, M. & Onnela, J. Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328**, 876–878 (2010).

[3] Ahn, Y., Bagrow, J. & Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **466**, 761–764 (2010).

[4] Evans, T. S. & Lambiotte, R. Line graphs, link partitions, and overlapping communities. *Phys. Rev. E* **80**, 016105 (2009).