Identity Lenses in Analyzing Evolving Social Structures

J. R. Hott[1], W. N. Martin[1], and K. Flake[2]

[1]Department of Computer Science, University of Virginia
[2]Department of Religious Studies, University of Virginia

In the effort to capture cultural dynamics, scholars have considered social networks, that is, a graph with people as nodes and their relationships as edges. These social networks are useful; however, to capture dynamics they must be considered over time. In the literature, Time-Varying Graphs (TVGs) [1, 3, 4] have been defined. In our investigations, we have found benefit in defining TVGs with nodes as social structures and people as the edges [5, 6] and then considering the dynamics of the social structures evidenced in the TVGs. Here we consider two motivating applications for our extensions to TVGs: early Mormon marital structures and an arXiv.org citation network.

The societal structures represented in the marital and church structures of early Mormons in mid-1800s Nauvoo, Illinois, include binary, polygynous, and polyandrous marriages, as well as child and adult adoptions, and membership of individuals in the church organization hierarchy. In this time period the concept of "marriage" is in flux and part of our research is to consider various conceptualizations of "marriage" to better understand the relationship to the formation of the church structure. Each conceptualization we consider as a different "identity lens," a term we create to describe these different views.

We therefore define the *identity-lens function* that maps one evolving network to another evolving network. More specifically, given a TVG $\mathcal{G} = (V, E, T, \rho, \zeta, \psi, \varphi)$ as defined in [4], the *identity-lens function* $f(i, \mathcal{G}) = \mathcal{G}_i$ maps the nodes and edges in $\mathcal{G}$ with a given identity definition $i \in I$ to a new Time/Identity-Varying Graph (TIVG) $\mathcal{G}_i = (V_i, E_i, T, \rho_i, \zeta_i, \psi_i, \varphi_i)$. $\mathcal{G}_i$ is therefore the view of $\mathcal{G}$ under identity lens i.

In our marital network research [5, 6], we represent marriages as the nodes, with the individuals connecting the marriages of their parents to their own marriages as adults. Every piece of this network is considered to be evolving, since marital relations change, new children are born, family members are adopted, and individuals change membership in the church organizational structure. Initially, this network $\mathcal{G} = \mathcal{G}_{binary}$ may be described as a binary-marriage network, in which each node depicts a marriage between two individuals and their biological children. This is one specific definition of node identity. However, we may examine this network in different levels of granularity: with different definitions of node identity. By extending the marriage definition to all individuals married to the same husband, we may redefine node identity to define polygynous marriages, creating $\mathcal{G}_{patriarchal}$. A related identity function that maps binary marriages to those with the same wife creates the TIVG $\mathcal{G}_{matriarchal}$[1].

[1]For those familiar with duals of graphs, $\mathcal{G}_{matriarchal}$ and $\mathcal{G}_{patriarchal}$ are not duals of one another. They are similar to each other in structure, but one may not be easily derived from the other, since a man may participate in multiple polyandrous marriages and a woman in many polygynous marriages.

Our second motivating example is the arXiv.org[2] citation network. ArXiv.org provides online open access to over 1 million cross-disciplinary papers, including papers in Physics, Mathematics, and Computer Science. We build a citation TVG $\mathcal{G}$ from this dataset, linking authors as nodes based on the co-authorship of their papers. Similar to the Nauvoo application, we define multiple identity functions to map this TVG to multiple TIVGs. Under a node identity function combining authors within the same institution, we produce $\mathcal{G}_{institutional}$. Other node identity mappings include departmental affiliation $\mathcal{G}_{departmental}$, and $\mathcal{G}_{subject}$, in which authors are mapped to their subject areas.

Each of these TIVGs have characteristics that change over time. As we increase the complexity of the nodes through the use of identity lenses, we increase the dynamics of the characteristics, specifically those captured within the nodes. In the Nauvoo dataset, these characteristics include familial relationships among marriage members and church leadership positions held by the members of each marriage. Similarly, in the arXiv dataset, the characteristics include departmental and institutional collaboration. We want and need metrics that are sensitive to these changes within the evolving nodes as well as the overall evolving structure of the network. To capture and analyze these dynamics, we first define flattening methods to produce static graphs depicting the state of the TIVG during a fixed-size interval around each time point, then compute centrality measures over the graph across time for each identity lens. This process creates a distribution of the metric across time, which may then be compared between identity lenses. We conjecture that utilizing different-sized flattening intervals and comparing distributions across identity lenses will provide insights to understanding the TVG and the motivating application it describes.

We therefore define two methods to flatten TIVG $\mathcal{G}_i$, in a λt-sized time interval around any given time-point $t \in T$, to a static graph $G_i(t, \lambda t) = (V_i(t, \lambda t), E_i(t, \lambda t))$. They are given by the following node and edge set definitions:

1. $V_i(t, \lambda t) = \{v \in V_i \mid \exists t' \in [t - \lambda t/2, t + \lambda t/2] \Rightarrow \psi_i(v, t') = 1\}$
   $E_i(t, \lambda t) = \{e \in E_i \mid \exists t' \in [t - \lambda t/2, t + \lambda t/2] \Rightarrow \rho_i(e, t') = 1\}$
   The union of all nodes and edges extant at any time during the interval[3].

2. $V_i(t, \lambda t) = \{v \in V_i \mid \forall t' \in [t - \lambda t/2, t + \lambda t/2] \Rightarrow \psi_i(v, t') = 1\}$
   $E_i(t, \lambda t) = \{e \in E_i \mid \forall t' \in [t - \lambda t/2, t + \lambda t/2] \Rightarrow \rho_i(e, t') = 1\}$
   Only nodes and edges that exist throughout the entire interval.

Using these flattening methods, we extend existing metrics to measure the dynamics of the graph across time. Let $G_i(t, \lambda t) = (V_i(t, \lambda t), E_i(t, \lambda t))$ be the flattened version of $\mathcal{G}_i$ in the λt interval around t using either of the methods above. Then, following the group definitions in [7], we define harmonic centrality [2], C_H, and betweenness centrality, C_B, of $G_i(t, \lambda t)$ as

$$C_H(G_i(t, \lambda t)) = \frac{\sum_{v_j \in V_i(t,\lambda t)} \left(\max_{\forall u \in V_i(t,\lambda t)} (C_H^*(u, G_i(t, \lambda t))) - C_H^*(v_j, G_i(t, \lambda t)) \right)}{[(|V_i(t, \lambda t)| - 1)(|V_i(t, \lambda t)| - 1)]/(2|V_i(t, \lambda t)| - 3)} \text{ and} \quad (1)$$

$$C_B(G_i(t, \lambda t)) = \frac{2 \sum_{v_j \in V_i(t,\lambda t)} \left(\max_{\forall u \in V_i(t,\lambda t)} (C_B^*(u, G_i(t, \lambda t))) - C_B^*(v_j, G_i(t, \lambda t)) \right)}{(|V_i(t, \lambda t)| - 1)^2 (|V_i(t, \lambda t)| - 2)}, \quad (2)$$

[2] `http://www.arxiv.org`

[3] As defined in [4], ψ_i and ρ_i are the "presence" functions for nodes and edges, respectively.

where $C_H^*(v_j, G_i(t,\lambda t))$ and $C_B^*(v_j, G_i(t,\lambda t))$ are the harmonic and betweenness centrality measures [7] for a given node $v_j \in V_i(t,\lambda t)$, respectively. The values of C_H^* and C_B^* rely on the flattening operation chosen, but the definition is agnostic, since it is based on $G_i(t,\lambda t)$. Without loss of generality, we will define only C_H^* here, using Boldi and Vigna's harmonic centrality definition [2], as

$$C_H^*(x, G_i(t,\lambda t)) = \sum_{\substack{d(y,x)<\infty \\ y\neq x\in V_i(t,\lambda t)}} \frac{1}{d(y,x)},$$

where $d(y,x)$ is the distance between $y, x \in G_i(t,\lambda t)$.

Allowing t to range over the entire lifespan of $\mathcal{G}_i$ and considering multiple sizes for our λt interval, we generate distributions of the metric across time and with differing levels of temporal granularity. These distributions give a picture of the dynamics occurring within each TIVG. By comparing the metrics across TIVGs under different identity functions for the same TVG, we hope to more fully capture the dynamics of and understand the original evolving network, and provide insights into the motivating application at hand.

We have therefore defined a new conceptualization of Time-Varying Graphs, specifically the identity-lens function and resulting Time/Identity-Varying Graphs under each identity mapping. We also defined methods for flattening the TIVGs into series of measurable static graphs and provided metrics over those representations. We intend to present the findings from these measures for each of our motivating applications: to better understand the definition of marriage in Nauvoo and its relation to church formation, and to illuminate patterns in author and departmental co-citations.

References

[1] Charu Aggarwal and Karthik Subbian. Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)*, 47(1):10, 2014.

[2] Paolo Boldi and Sebastiano Vigna. Axioms for centrality. *Internet Mathematics*, 10(3-4):222–262, 2014.

[3] Arnaud Casteigts, Paola Flocchini, Emmanuel Godard, Nicola Santoro, and Masafumi Yamashita. Expressivity of time-varying graphs. In *Fundamentals of Computation Theory*, pages 95–106. Springer, 2013.

[4] Arnaud Casteigts, Paola Flocchini, Walter Quattrociocchi, and Nicola Santoro. Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, 27(5):387–408, 2012.

[5] John R Hott, Worthy N Martin, and Kathleen Flake. Evolving social structures: Networks with people as the edges. *Digital Humanities Forum, University of Kansas*, 2014.

[6] John R Hott, Worthy N Martin, and Kathleen Flake. Visualizing and analyzing identity classes in evolving social structures. *Chicago Colloquium on Digital Humanities and Computer Science*, 2015.

[7] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*, volume 8. Cambridge University Press, 1994.