2023

Bayesian Social Subgraph Generative Models: Social Network Twins using Belief Networks and Ego Behavior Models

Nagraj Naidu

Omar El-Gayar

Bayesian Social Subgraph Generative Models: Social Network Twins using Belief Networks and Ego Behavior Models

Nagraj Naidu
Dakota State University
nagraj.naidu@trojans.dsu.edu

Omar El-Gayar
Dakota State University
omar.el-gayar@dsu.edu

Abstract

A key assumption of a Subgraph Generative Model (SUGM) for sparse networks is that a subgraph is independent of lower order subgraphs in a sparse network. This is not entirely true especially for non-sparse networks. Additionally, the generated networks lack the typical properties of a social network because of an assumption of random growth for nodes and edges. Finally, there is no concept for explicit ego choice or bias when connecting to dyadic or triadic relationships. We develop a novel graph generative model referred to as the Bayesian Social Subgraph Generative Model (BASSUGM). We ground the BASSUGM in a proposed sociological model and leverage Bayesian tools like belief networks. We introduce novel concepts like the networks' macro theme when combines with an ego's individuality realizes the ego's intent. We also demonstrate how the social network twin generated with BASSUGM outperforms SUGM for non-sparse, small, social, networks.

Keywords: Generative Model, Graph Theory, Bayesian Belief Network, Behavior Model, Network Twin

1. Introduction

A social network is an embodiment of social life between people. It reflects the positions they hold, roles they play and the relationships they form or break. People help, hinder, trade, fight, and be-friend. In the social network of relationships people realize their desires and express their intent. Social networks are an "invisible structure that underlies society and has its influence in determining the conduct of society as a whole" (Moreno, 1993).

Sociological formations are characterized by simple structures and their interactions (Wolff & Simmel, 1950). The simplest structure is the isolated individual who has no interaction with the network. Simmel defines an isolated individual as a temporal and interrupting social relationship. "Isolation is a relation which is lodged within an individual, but which exists between him and a certain group". (Wolff & Simmel, 1950)

A network twin is a representation of an existing real-world network. We need network twins to model dynamic, interacting, social, and temporal real-world phenomenon. The use cases for network-twins range across descriptive, predictive, and prescriptive analytics. These use cases also span a gamut of domains include internet of things (IOT), fraud, community mining, and customer churn.

In specific cases, the practical applications of twin networks require high representational fidelity in terms of the networks actor's behavior and relationships. This is especially true for critical applications like customer fraud detection, customer churn management, or anti-terrorism financing. Low fidelity or coarse network representations that focus only on global properties are of limited value. This is because the focus of descriptive, predictive, and prescriptive analytics in modern network applications is not only on the global network neighborhood but often the actors local, immediate neighborhood.

The overall problem under consideration is the discovery of a straightforward way to construct network twins with minimal loss of structural and behavioral fidelity. The Subgraph Generative model (SUGM) proposes generation of a digital network twin for large sparse graphs in a straightforward way using a cumulative generation and overlay of subgraphs. First, it generates nodes. Then it connects links or dyads. Finally, it connects triangles or triads. However, the SUGM makes certain assumptions. A central assumption is the subgraph independence for large sparse networks. This becomes a problem for non-sparse networks where there is a more pronounced dependence between subgraphs. Another assumption is the global and random nature of subgraph connectivity. This is also a problem because in addition to connecting using a global connectivity principle, nodes also connect using a local connectivity mechanism that reflects the nodes local choices and bias. Hence, the decision to connect two or more egos is a combination of global trends and local choice. This is important because we want the digital network twin structure to reflect sociological behaviors of nodes in the network as well as the macro connectivity theme.

In this research paper propose a novel method to generate digital network twins for social networks that is simple to implement and still provides the structural and behavioral fidelity in the network twin. Our proposed solution uses Bayesian belief networks to incorporate conditional subgraph dependence as observed in the original social network. We use this conditional subgraph dependence to probabilistically generate subgraphs without the assumption of independence. Additionally, our proposed solution combines macro concerns with local behavior to generate probabilities for connectivity. We call our network generation process the Bayesian Social Subgraph Generative Model (BASSUGM).

We structure the sections in the research paper as follows. In Section 2, we discuss the sociological model of a network. In Section 2.1 we discuss prior work on graph generative models in depth. In Section 3, we describe our approach to operationalize our network generation process in detail. We include the details of the probabilistic belief network as well as the dyadic and triadic behavior models which we use for the ego's bias. In Section 4, we evaluate our proposed generative model by forming hypotheses and experimenting with popular non sparse, and small social networks. In Section 5, we discuss and reflect on the experiment results. In Section 6, we discuss the potential for future research work. Finally, in Section 7, we conclude the paper.

2. Background and related work

An actor is a node in a social network. Ego is a focal actor. The ego is defined as "the self especially as contrasted with another self or the world" ("Ego.," 2022). Ego nodes connect to other ego's called alters. The alter is defined as "a second self or different version of oneself" or as "a trusted friend" ("Alter Ego.," 2022).

A relation is a tie or link between an ego and an alter. An ego can have multiple relationships in the network. We will call the set of ego relationships as a role. (Lorrain & White, 1971) describes role as "consisting of sets of relations linking this person as ego to sets of others". Roles are local and are a characteristic of the ego. In this paper, we use the term degrees from graph theory to denote the count of ego relations.

An ego is associated with a position in the network. There are a set of egos who share equivalent role patterns. Egos with similar positions are structurally equivalent.

Egos are free. To be free is to have choice. To have choice is to realize (make concrete) an individual's intent to relate to other alters in the network.

A dyad is the first structure where there is relationship between two egos. It is the germ of the society where the individuality of the members is favored equally (Wolff & Simmel, 1950). The dyad is unique in that either member has equal power to dissolve the dyad.

A triad expands a dyad and brings a third to the relationship. There is a direct relationship between two individuals but mediated by the indirect one of the third. This leads to a dynamism of the relationship. Power and the strengths of relationships is unequal and shifting in a triad. The power relationship in a triad contributes to the temporal dynamism of the network and is an important consideration in network construction.

However, the full realization of the ego's intent to create or maintain a relationship does not depend entirely on the ego's choice. There are global influences at play that moderate the full realization of the intent. We call these global influences, the macro theme.

Macro themes apply to the entire network or a portion of it. This is the invisible structure of networks characterized by (Moreno, 1993). Macro themes are the fabric of the network and moderate the realization of individual choices. The realization of the ego's intent is an outcome contributed by both the macro theme and the ego's intent. For example, an individual (ego) wishes to stay connected with a school friend (alter), but macro considerations like the distance, boundaries like weather, or even events like war moderate the intent. Note the reference to "invisible structures" in (Moreno, 1993). The full realization of an individual's intent depends in part to those invisible structures which are beyond the individual's local intent and indeed, control. The individual "ties others and is tied by others" (Wolff & Simmel, 1950). The invisible structures that influence the ego's intent and indeed the network formation are the macro themes of the social network.

The sociological model of the network that we have described above is important because it conceptualizes and guides our network generation process. We operationalize this conceptual model in Section 3.

2.1. Graph generative models

The goal of graph generative models is to construct a representation of an observed network (Wasserman & Pattison, 1996). We can categorize graph generative models as feature driven, structure driven, and intent driven. (Lim et al., 2016). Feature driven models define a "mechanism or a principle by which a network with desired features is constructed." One way to connect networks is to probabilistically connect network nodes and edges using a uniform distribution (Erdös & Alfréd, 2011). Another way is to use a preferential attachment model where the node attachment probabilities are proportionate to the number of node ties. (Barabási & Oltvai, 2004). Structure driven models capture global properties to generate the network. dK-Graph (Mahadevan et al., 2006) measures and generates random graphs by capturing probability distributions of the subgraph's properties. Intent driven models emulate actor relationships. Random walks (Vazquez, 2003) probabilistically create nodes and ties while traversing the network.

Hybrid networks combine dynamic random and dynamic preferential attachment networks. (Jackson, 2008). The rationale is that networks in practice are a proportional combination of the preferential attachment and random networks.

The U|MAN distribution is a uniform distribution conditioned on the dyad census and includes mutual dyads, asymmetric dyads and null (or unconnected) dyads <M,A,D> (P. Holland & Leinhardt, 1974). The distribution emits the conditional probabilities of node connections

The "p1" model are four log-linear models whose outcomes were the probabilities of the <M, A, D> dyad census (P. W. Holland & Leinhardt, 1981). One can introduce actor attributes into log-linear models. However, these models impose severe independence assumptions. The p* models also called Exponential Random Graph Models (ERGM) is a linear combination of coefficients and network count statistics (Fienberg & Wasserman, 1981). The aim is to find the probability of a network observation over all networks and then use the estimated coefficients to characterize the network. The estimation uses Markov Chain Monte Carlo (MCMC) methods to sample the networks. One difficulty is that it is hard to get stable parameter estimates for a large sparse network because the number of all subgraphs is prohibitively large for sampling.

Subgraph Generation Models (SUGM) is a layered and incremental subgraph generative approach (Chandrasekhar & Jackson, 2016, 2021). In the SUGM, there is an incremental generation of subgraphs. These subgraphs are overlayed on top of lower order subgraphs. The motivation for SUGM was the difficulty in parameter estimation in ERGM. In SUGM, dynamic network growth builds the network using subgraphs.

A key assumption of a SUGM is that a subgraph is independent of lower order subgraphs in a sparse network. As an example, we could assume that a triangle is independent of links. However, triangles are not independent of links. A simple illustration in which removing a link removes one or more triangles, proves this point. So, the probability estimations as calculated in the SUGM will not capture any lower order subgraph dependence. Additionally, the generated networks lack the typical properties of a social network because of random network growth. Typical social networks properties include fat tails, small diameters, scale free behavior, small average path lengths and assortativity. Finally, there is no consideration of the ego's choice. The ego has a preference bias when choosing an alter in a dyad or a pair of alters in a triad.

There are two structural limitations in the SUGM. One, subgraphs may be incidental. This means that the observed count of the subgraphs may not be accurate. This is important because for network twins we want the subgraph counts of the network twin to be close to the original subgraph as possible and to get the closest fit to the original network. Two, subgraphs are not independent. These difficulties may not be severe for large sparse networks. However, this difficulty impacts smaller networks. This is important because we want to capture the dependence nature of the original network in the network twin.

Additionally, as previously discussed the ego exercises choice to realize their intent. However, the network macro theme moderates this choice. The inclusion of explicit behavior and this combination of behaviors is missing in the SUGM and other generative models in the literature. By explicitly including these behavior models we can probabilistically connect nodes not just on a global theme like preferential attachment but also on local choices of the nodes or egos.

The BASSUGM approach uses probabilities of nodes, links and triangles conditioned on lower order subgraphs to consider the dependent nature of networks. Furthermore, this approach uses local behavior models that consider the nodes choice of association and combine these local choices with the global choice characteristics of the network.

We use graph theoretic notation in this paper. To express probabilities, we use common notations from probability theory. Finally, to describe causal Bayesian networks we use directed influence diagrams and conditional probability tables. We also use plate notation to represent variables in a Bayesian regression.

3. The approach

3.1. Bayesian social subgraph generated model

The aim of the BASSUGM is to construct a close representation of an observed network. It operationalizes key concepts in Section 2. Table 1 maps the key concepts.

To summarize the procedure, the BASSUGM replaces the estimated probabilities of a SUGM with the conditional joint distribution of the subgraph over all lower order subgraphs using a Bayesian belief network. To include social network like properties in the generated network, the BASSUGM proposes preferential attachment that simulates the small world behavior. To account for dyadic local choice, the BASSUGM uses a dyadic local behavior model. The models are logit models and provides the probability of an ego node "choosing" an alter node. In subsequent sections we argue the applicability of the logit model. This probability of dyadic choice combines with the probability of the preferential attachment to account for the macro theme and the full realization of the ego's intent. Finally, BASSUGM also considers the ego's triadic choices. Triads are added or removed by adjusting the triad count at random like recommended for the SUGM in (Chandrasekhar & Jackson, 2016). In addition, in the BASSUGM, we also include the propensity of triangle formation from a triadic local behavior model.

Table 1. Sociological concept mapping

Sociological Concept	Operationalization Mapping
Ego	Section 3.3: Ego Generation
Ego's Position and Role	Section 3.4: Ego Position Generation
Network Hidden Structures (Macro Theme)	Section 3.5: Dyad Intent Generation
Ego's Dyadic Intent	Section 3.5: Dyad Intent Generation
Combine Hidden Structures and Dyad Intent	Section 3.6: Combine Macro Theme and Dyad Intent
Ego's Triadic Intent	Section 3.7: Triad Intent Generation

3.2. Subgraph causal belief network

In the BASSUGM, we use the discrete condition probabilities distribution (CPD) of each subgraph as a node over the lower order parent in a causal belief network. We want to establish the joint distribution P (Ego, Link, Triangle) and infer conditional probabilities over the joint distribution. The belief network assumes local independencies. As an example, triads are independent of ego given its parent. i.e., Triangle ⊥ Ego | Link.

We define the CPD for each node of the belief network. Egos have no parent but have two states – An isolated or a non-isolated state. Ego is the parent of Link. Links states are the unique counts of observed dyads for all egos in the network conditioned on ego states. Link is the parent of triangle conditioned on link states. Triangle states are the unique counts of all observed triangles for all egos in the network.

The Bayesian network computes the joint distribution of the network using the chain rule of probability and applying the local independence condition. To compute conditional probabilities, we use Bayesian software for variable elimination to marginalize over all other variables.

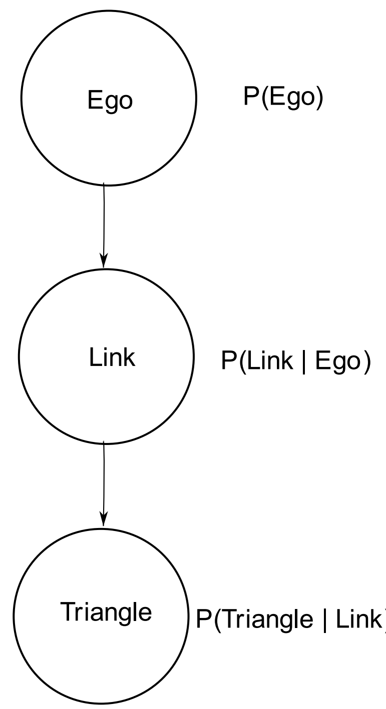

Figure 1. Subgraph causal belief network

We define the CPD for each node of the belief network. Egos have no parent but have two states – An isolated or a non-isolated state. Ego is the parent of Link. Links states are the unique counts of observed dyads for all egos in the network conditioned on ego states. Link is the parent of triangle conditioned on link states. Triangle states are the unique counts of all observed triangles for all egos in the network.

The Bayesian network computes the joint distribution of the network using the chain rule of probability and applying the local independence condition. To compute conditional probabilities, we use Bayesian software for variable elimination to marginalize over all other variables.

Table 2. Ego conditional probability distribution

State	
Isolates	P (Ego | Isolates)
Non-isolates	P (Ego | non-Isolates)

Table 3. Link conditional probability distribution

	Evidence	
State	non-isolates	isolates
Degree = 0	P (Degree = 0 | non-isolates)	0
Degree = 1	P (Degree = 1 | non-isolates)	0
Degree = n	P (Degree = n | non-isolates)	0

Table 4. Triangle conditional probability distribution

	Evidence		
State	Degree = 0	:	Degree = k
Triangle Count = 0	P (Triangle Count = 0 | Degree = 0)	:	P (Triangle Count = 0 | Degree = k)
:	:	:	:
Triangle Count = n	P (Triangle Count = n | Degree = 0)	:	P (Triangle Count = n | Degree = k)

3.3. Ego generation

In BASSUGM, like in the SUGM we generate "nicely ordered" subgraphs starting from an empty network. First, we create a seed network from a fixed number of nodes. We choose the star network as the seed because it best represents a basic dyadic formation of an ego actor connected to multiple alter actors. We then introduce egos dynamically over time. Time t generates a node n, then at time t + 1 a new node n+1 generates. By randomly choosing from a categorical probability distribution with two categories, isolates, and non-isolates, we tag the ego as isolate, or non-isolate. We infer this categorical distribution from the Bayesian belief network.

Table 5. Ego categorical distribution

Parameter	Description
Number of Categories (k)	k = 2 {Isolates, non-isolates}
P(Ego)	$\{p_1, p_2\}$ where $p_i >= 0$, $\Sigma\, p_i = 1$

3.4. Ego position generation

We now generate ego positions. Recall that an ego position is simply the count of relations for the ego. We randomly choose the count of relations from a categorical distribution where the categories are a sequence from zero up to the maximum relation count in the observed network. We infer the categorical distribution from the belief network.

Table 6. Position (Links) categorical distribution

Parameter	Description
Number of Categories (k)	n = maximum of ego degrees $\{k_0, k_1, \ldots, k_n\}$
P(Links|Egos)	$\{p_1, p_2, \ldots, p_n\}$ where $p_i >= 0$, $\Sigma\, p_i = 1$

3.5. Dyad intent generation

As discussed in Section 2, the realization of the ego's intent is a relation to which, the macro theme and the local role contribute.

Preferential attachment approximates the macro theme in a social network. We use this as our macro theme. To choose an alter, we prepare a categorical probability distribution proportional to the number of degrees of all existing egos in the network. We then can choose randomly to get the dyadic alter for the ego.

We want to account for local behavior. To model this, we use a dyadic local behavior model. The model is a logistic random effects model that generates the score of a relation between the ego and alter. The inputs of the model are structural components of the nodes. The outcome is a Bernoulli variable. We prepare a vector whose elements are the egos in the network and the values are the scores from the dyadic local behavior model. The vector represents the score of all possible alters for an ego.

Table 7. Preferential attachment categorical distribution

Parameter	Description
Number of Categories (n)	n = count of egos n > 0
Probability Distribution Vector (V)	V = {p_1, p_2, p_n} where $p_i >= 0$, $\Sigma p_i = 1$ d_i= degrees of ego i $p_i = d_i / \Sigma d_i$

We justify the use of logistic regression for the behavior model as follows. Dyadic Interaction Models in the literature use log linear generalized linear models (GLM) called the p1 model. If we assume relationships are dichotomous. i.e., are present (1) or absent (0), it translates into a Bernoulli distribution i.e., k successes out of n observations.

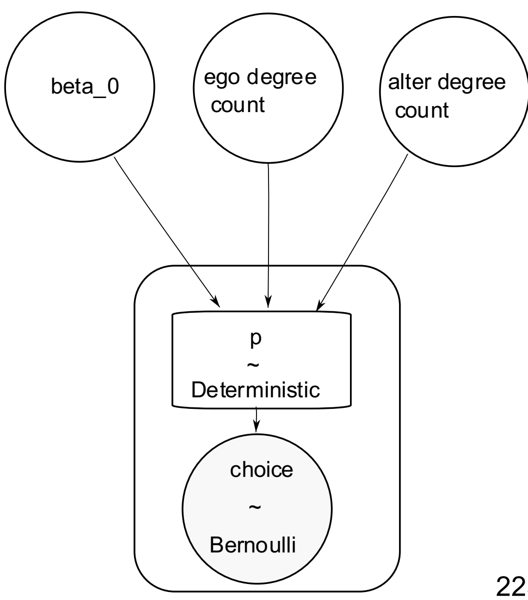

Figure 2. Dyadic local behavior model

We can then assume a logistic model whose dependent variable is a Bernoulli random variable. A central assumption of the logistic model is that observations are independent. For the Dyadic Local Behavior Model, we can argue that the basic modelling unit is the Dyad, which means that all observations are replicable. We also use the count of relations (degrees) for the ego and alter for independent variables. We do not use compositional attributes that involve dependencies. With this assumption, we can justify the use of logistic random models.

3.6. Combining macro theme and dyad intent

To completely realize the ego's intent, we must combine the outcomes of the preferential attachment and the local behavior model. We assume that both probabilities are in a set of pairwise disjoint events whose union is the entire sample space.

Bi = {macro theme event, local choice event}

If event A is the realization of the ego's intent, then by the law of total probability:

P(A) = Σ P (A | Bi) P(Bi)

If we assume a discrete uniform distribution for the prior of both events, this then translates to a simple unweighted average of probabilities. We can now take a simple average of the preferential categorical distribution with the normalized dyadic behavior scores vector to get a new categorical distribution. We randomly select the alter from this new distribution.

3.7. Triad intent generation

At this point we have grown the network and have generated egos and dyads. We still need to include triads.

Dyad formation leads to accidental or incidental triads. Two dyads may very well form a triad just by chance and not explicit intent. For our representational network to be as close to the original network as possible, number of triangles in the network must closely match the number of triangles in the observed network. We use a variation of the proposed approach in (Chandrasekhar & Jackson, 2016). In general, we remove or add new triangles to the network until we approximate the observed number of triangles. To get the expected number of triangles for a node we probabilistically pick the number of triangles from a categorical distribution of count of triangles conditioned on the number of node links from the belief network.

We randomly select an ego, randomly select two alters, and then add or remove their relation to update the triangle count. We continue the process until we obtain the expected triangle count.

Table 8. Triangle categorical distribution conditioned on link count

Parameter	Description
Number of Categories (k)	n = number of triangles {0, 1, …., n}
P (Triangle | Link = m)	n = maximum of all unique relations {k_0, k_1, …., k_n} m = count of ego degrees

However, a random selection of two existing alters does not take into consideration local choice. To improve the random selection of a connection and to model the local triadic choice, we use a triadic local behavior model. The model is a logistic random effects model that emits the propensity score of a relation between the ego and both alters. The inputs of the model are structural components of the actors. The outcome is a Bernoulli variable.

To select a relation between two alters for addition or deletion, we score all combinations of alters for the ego. For deletion we target the pair with the lowest score. Conversely for addition, we target the pair with the highest score. We can interpret this as the ego de-friending their least compatible friend or friending the alter with the highest friend propensity.

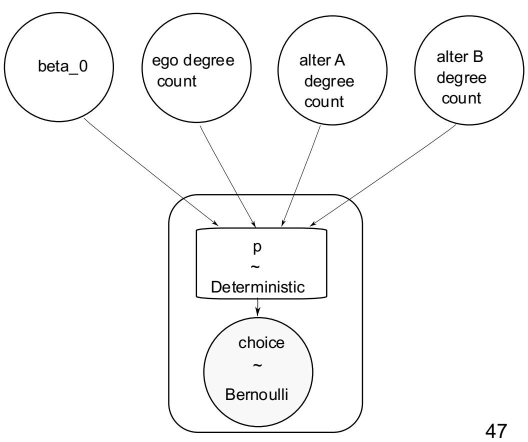

Figure 3. Triadic local behavior model

4. Evaluation

We use three well-known social networks as benchmarks for our experiments:

- Padgett's Florentine Families (FF)
- Zachary's Karate Club (KC)
- Knuth's Les Misérables (LM)

For the purposes of this research, we treat all networks as non-directional, one mode networks, and single relation. We execute our implementation of SUGM and BASSUGM on the three benchmark networks. The SUGM implementation follows the algorithm proposed in (Chandrasekhar & Jackson, 2021) for small networks. For the generation of dyads, we vary preferential attachment, random selection, and preferential attachment with the dyadic local behavior model. For triangle generation, we vary random selection and the triadic local behavior model.

We use the following measures to compare the observed and the twin networks.

- Average Degree
- Triangle Count
- Average Clustering
- Global Transitivity
- Global Density

Table 9. Benchmark network statistics

	Nodes	Edges	Mean Degree	Global Clustering Coefficient	Sparsity Index
FF	16	35	4.38	0.30	0.709
KC	34	77	4.53	0.26	0.863
LM	77	254	6.60	0.5	0.914

The measures of both the observed and generated networks are vectorized after normalization. We then measure the cosine similarity of the two vectors. BASSUGM generates a random network. The generated network is just one random choice from the set of all networks. To ensure reliability we take the average of the measures from thirty random generated representations for our comparisons. We assume that 30 samples are a boundary for a large sample which is a popular statistical rule of thumb. We calculate the cosine similarity after link and triangle generation. We use the metrics to evaluate the following hypotheses:

H1: BASSUGM performs better than SUGM when generating representations for small non-sparse social networks.
H2: Use of the dyadic local behavior model to select an alter in BASSUGM generates a better representation as compared to using SUGM.
H3: Use of the triadic local behavior model to select triangles in BASSUGM generates a better representation as compared to using SUGM.

In general, the FF, KC and LM networks get sparser. We use a sparsity measure that is one minus the edge density. We can also express this in the following formula for the sparsity index, where n is number of nodes and e is number of edges.

$$sparsity\ index = 1 - \frac{e}{\binom{n}{2}}$$

We first generate benchmark representational networks using the SUGM. Note that the performance of the SUGM as measured by the similarity measure degrades with the increase in the network global clustering coefficient. As network clustering and sparsity increases, dependencies among dyads and triads increase and the assumption of independence for isolates, links, and triangles in the SUGM becomes increasingly invalid. Next, we use three BASSUGM configurations that correspond to our three hypotheses. In the first configuration, we generate links using preferential attachment and then update triangles by selecting alters randomly. This is the basic BASSUGM configuration and the first building block for other configurations. In the second BASSUGM configuration, we build on the first configuration and generate links by using preferential attachment and a dyadic behavior model which injects the nodes local choice. Finally, in the third configuration, we build on the second configuration by using a triadic behavior model to inject the nodes triadic choice.

5. Results and discussion

Table 10 summarizes the results of the experiments. Generating the representational network using the first configuration we find that the BASSUGM outperforms SUGM for all three networks. In addition, the second configuration BASSUGM also outperforms the SUGM for one of the networks (LM) and ties for the other two networks. For the third configuration BASSUGM ties for the LM network and does not outperform the SUGM for the other two networks. Thus, we fail to reject our first hypothesis H1 and conclude that BASSUGM outperforms SUGM in at least one configuration of the BASSUGM.

The second configuration extends the first configuration by using a dyadic behavior model to account for the node's local choice. This configuration outperforms one network (LM) and ties for the other two networks. Note that the BASSUGM outperformed the SUGM in the network with the highest global clustering coefficient and the highest sparsity. These types of networks consist of clusters that tightly connect to each other and do not tightly connect to nodes outside the group. In these networks local choice has a greater impact than the global theme. By using the dyadic behavior model, we inject this behavior explicitly during network construction which results in a better network representation. Hence, the results support hypothesis H2 and we can conclude that using the dyadic local behavior model in BASSUGM generates a better representation as compared to using SUGM especially for networks with high sparsity and high clustering coefficients which are typical for large social networks.

The third configuration extends the second configuration by adding a triadic behavior model. This configuration does not outperform the SUGM or any other BASSUGM configuration. Hence, the results do not support hypothesis H3. One reason could be that we did not include shared attributes of the triad actors in the regression. Another reason could be the inherent dependent nature of the triad which increases with the increase in clustering. We will include shared ego attributes and analyze dependencies for triadic behavior models in future research.

Overall, increasing the sparsity of the network resulted in a decrease the performance of BASSUGM for small networks. As sparsity increases the global probabilities or the global theme become less relevant and local behavior becomes more pronounced. We used unweighted priors while combining macro and local behaviors. Weighing priors proportionately to the sparsity index may better similarity performance but we have postponed this reasoning for future research.

Table 10. Experiment results *

	FF	KC	LM
SUGM			
	0.95	0.95	0.86
Configuration 1: BASSUGM Generate links with preferential attachment + Select triangle alters randomly			
	0.96	0.96	0.91
Configuration 2: BASSUGM Generate links with preferential attachment and the dyadic behavior model + Select triangle alters randomly			
	0.95	0.95	0.90
Configuration 3: BASSUGM Generate links with preferential attachment and the dyadic behavior model + Select triangle alters with the triadic model			
	0.93	0.94	0.86

*Cosine similarity of evaluation vectors

The BASSUGM is especially useful in applications that need descriptive, predictive, and prescriptive analytics at the actor level. For example, we can construct a telephone network that needs a churn analysis application using the BASSUGM process. In this type of network, customers are nodes, and the call connections are relationships. Initially, we can create new isolate customers. Then, we can establish the number of connections (position). The telephone network is social so it would use preferential attachment to connect the nodes as a first step. This will form the fabric of the network. Then the dyadic behavior model can generate the appropriate customer connections at a local i.e., per customer level. After which the triadic behavior model could connect the triadic relationships for each customer.

6. Future work

There are multiple ways to extend this research. First, the observed network is ever evolving. Egos join the network and leave. Macro themes are temporal and constantly changing. Ego behaviors will vary across groups and will change over time. How will the representation network synchronize with changes in the observed network? This is a key research area because of the real time use of network twins in analytics and predictions.

Second, the dyadic and triadic local behavior models use "structural properties." These are properties that involve pairs of actors like the count of relationships. One class of information are compositional variables, which are attributes attached to the actor like gender. How can we incorporate actor attributes in these models?

Third, there are other types of more complex networks like multigraphs (more than one actor relations) and hypergraphs (affiliation networks). In directed graphs, relations are directional which adds additional sociological concepts. A good example is marriage. Marriage is a directed relationships either way from two actors. Future extensions to this work may take into consideration these additional concepts. One is reciprocity which is the strength of the choice or the exchange of the directed relationship. In the marriage example above, the strength of the tie can be an important consideration of a stability prediction. Bipartite graphs are popular in practice because they include concept abstraction between actors. We can add these concept abstractions. For example, in a corporation network, corporations share resources. These lead to bipartite graphs because there is resource sharing between corporations. Thus, a resources node connects two corporation nodes. Finally, temporal concepts are important in network applications. We can include temporal concepts including time and the changes to network structure over time.

Finally, modern social networks are large with millions of nodes. It is difficult or impossible to directly prepare representations because of the scale and complexity of the networks. Usually, the broad answer to this is sampling. However, sampling should capture both macro and behavioral details of the network and go beyond capturing only the rough structural representation.

7. Conclusion

The proposed model conceptualizes social networks using Bayesian techniques. It accounts for concepts like an ego's behavior and its freedom to connect to their choice of an alter. Furthermore, it makes a distinction between macro trends in the network that invisibly affect an ego's choice and local behavior that asserts the ego's individuality. The result is a network representation that is closer to sociological reality. We use the union of subgraphs in the SUGM as foundation and then enhance the BASSUGM with belief networks and logistic behavior models. We show that BASSUGM social network twins outperform SUGM generated representations.

In conclusion, network generation is not merely about the generation of the best statistical representation of an observed network. Instead, it is paramount to be able to incorporate important sociological behaviors and other concepts to get the most useful representation for the myriad of use cases that network representation has in analytics, forecasting, and simulation.

8. References

Alter ego. (2022). In *Merriam-Webster.* https://www.merriam-webster.com/dictionary/alter%20ego

Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. https://doi.org/10.1038/nrg1272

Chandrasekhar, A. G., & Jackson, M. (2016). *Tractable and Consistent Random Graph Models*. 61.

Chandrasekhar, A. G., & Jackson, M. O. (2021). A Network Formation Model Based on Subgraphs. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2660381

Ego. (2022). In *Merriam-Webster.* https://www.merriam-webster.com/dictionary/ego

Erdös, P., & Alfréd, R. (2011). On random graphs. *Publ. Math. Debrecen*. https://www.degruyter.com/document/doi/10.1515/9781400841356.38/html?lang=en

Fienberg, S. E., & Wasserman, S. (1981). *Categorical Data Analysis of Single Sociometric Relations*. 57.

Holland, P., & Leinhardt, S. (1974). The Statistical Analysis of Local Structure in Social Networks. *National Bureau of Economic Research, Inc, NBER Working Papers*, *6*.

Holland, P. W., & Leinhardt, S. (1981). An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, *76*(373), 33–50. https://doi.org/10.1080/01621459.1981.10477598

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.

Lim, S.-H., Lee, S., Powers, S. S., Shankar, M., & Imam, N. (2016). *Survey of Approaches to Generate Realistic Synthetic Graphs* (ORNL/TM--2016/3, 1339361; p. ORNL/TM--2016/3, 1339361). https://doi.org/10.2172/1339361

Lorrain, F., & White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of Mathematical Sociology*, *1*(1), 49–80. https://doi.org/10.1080/0022250X.1971.9989788

Mahadevan, P., Krioukov, D., Fall, K., & Vahdat, A. (2006). Systematic Topology Analysis and Generation Using Degree Correlations. *ACM SIGCOMM Computer Communication Review*, *36*(4), 135–146. https://doi.org/10.1145/1151659.1159930

Moreno, J. L. (1993). Emotions Mapped by New Geography. *New York Times*.

Vazquez, A. (2003). Growing networks with local rules: Preferential attachment, clustering hierarchy and degree correlations. *Physical Review E*, *67*(5), 056104. https://doi.org/10.1103/PhysRevE.67.056104

Wolff, K. H., & Simmel, G. (1950). *The Sociology of Georg Simmel.* https://archive.org/details/sociologyofgeorg030082mbp/page/n35/mode/2up