Mantzaris *et al. Comput Soc Netw* (2021) 8:12
https://doi.org/10.1186/s40649-021-00095-y

Computational Social Networks

RESEARCH Open Access

Utilizing the simple graph convolutional neural network as a model for simulating influence spread in networks

Alexander V. Mantzaris[1*], Douglas Chiodini[1] and Kyle Ricketson[2]

*Correspondence:
alexander.mantzaris@ucf.edu
[1] Department of Statistics and Data Science, University of Central Florida (UCF), 4000 Central Florida Blvd, Orlando 32816, USA
Full list of author information is available at the end of the article

Abstract

The ability for people and organizations to connect in the digital age has allowed the growth of networks that cover an increasing proportion of human interactions. The research community investigating networks asks a range of questions such as which participants are most central, and which community label to apply to each member. This paper deals with the question on how to label nodes based on the features (attributes) they contain, and then how to model the changes in the label assignments based on the influence they produce and receive in their networked neighborhood. The methodological approach applies the simple graph convolutional neural network in a novel setting. Primarily that it can be used not only for label classification, but also for modeling the spread of the influence of nodes in the neighborhoods based on the length of the walks considered. This is done by noticing a common feature in the formulations in methods that describe information diffusion which rely upon adjacency matrix powers and that of graph neural networks. Examples are provided to demonstrate the ability for this model to aggregate feature information from nodes based on a parameter regulating the range of node influence which can simulate a process of exchanges in a manner which bypasses computationally intensive stochastic simulations.

Keywords: Graph convolutional neural networks, Social influence, Networks, Machine learning, Social networks, Information diffusion

Introduction

The study of networks [1] has shown that they are ubiquitous in nature and that many complex processes that are studied [2] have behaviors defined by their network structure. The famous initial formulation of a problem where a network (graph structure) is explicitly defined is in the *Seven Bridges of Königsberg problem* [3] which brought questions for which graph theory and topology later took inspiration from, growing into fields of their own. The question of how to traverse nodes in a network in an optimal manner is still an active area of research commonly referred to as the *travelling salesman problem* [4]. It is worthy to note that these networks were noticed historically examining the human built structures we live within, but network founded organization principles have been in use for millions of years in metabolic cycles [5] and within food webs [6]. As

society has been building networks, and relying upon other networks to support it, the number of nodes is increasing as well as the clustering coefficient for those network sizes [7] (Fig. 1 shows this for biological and non-biological network data). The observation of how these fundamental networks persist over time allows one to speculate that online social networks such as Facebook, Twitter, and Instagram, for instance, will also remain as a large component of our interconnected society in some form, as stated in [8]. The question explored in this paper is how to model node label classifications based on the combination of the features each node contains, and the incorporation of the features held by nodes in the vicinity which changes size resulting in different accumulations of *local influence*. Specifically, it addresses how the number of hops taken into consideration in a graph neural network can change the node label predictions due to the range of influence in the vicinity. This is important when attempting to provide categorical labels to nodes representing users based on their attributes and associations.

Simplifying these networks with many unique node IDs into a smaller set of labels assists in the effort to simplify the data and generalize with respect to aggregated behaviors. An example is with voting patterns where each eligible voter may have a unique perspective on the issues determining a choice, but, ultimately, this information is pigeonholed into a smaller set of expressible options where many will have their choices coalesced with others of different yet similar opinion holders. This concept is applied in collaborative filtering [9] typically for retail where recommendation systems use a customer's interests with that of a community set to predict an affinity for new items. There can be a set of communities which a potential customer can be compared with to find the maximum expected overlap. What this does not take into account is whether the person has links (edges) with a specific set of nodes in those communities, and how that information can be useful for the predictions. Nodes can be expected to produce actions based on the information contained in the features and the edges where an ideal label allocation relies on both sources of information. The value of the edge set inclusion can bring to mind the colloquial phrase 'birds of a feather flock together' [10] which highlights how *homophily* based link creation is prevalent in human populations for a wide range of reasons, and this is displayed in the social networks produced through local interactions, [11]. As a consequence, the edge set can drive inferences of changes in the label allocation even if the inferences of an optimal feature distance metric says otherwise. How this phenomenon of homophilic edge creation arises in networks during a growth phase is very interesting, [12, 13], but in this work, we consider that the link set expressing the homophily is provided. These links may be defined by platform-dependent actions such as 'friendship' or 'following' links, and the features can be a set of variables categorical or numerical in nature.

In the effort to understand how the features of a node can propagate through a network to influence a node's categorization, insight can be drawn from the research of information spread (or information diffusion [14]), as pieces of content are shared between users/agents within online social networks. The established Katz centrality measure [15] is based on the number of 'walks' [16] with a penalization for the number of edges traversed in a walk between nodes as a measure for centrality and communicability (in static and temporal networks [17, 18]). A key concept is that there is a ranking in the ability for nodes to spread information between each other that is inversely proportional

to the number of edges traversed [1]. These measures are useful in marketing applications where brands seek to exploit the sharing mechanism provided by the platforms for users themselves to amplify visibility of their marketing campaigns [19] (e.g., 'growth hacking'). With various trends emerging from the sharing of content, it then becomes a question of where to place the credit for the spread; to the initiation with the crowd or to opinion leaders as explored in [20]. It is important to note that these approaches are not asking whether features of one node affect another node's general behavior, since the content transmission can be irrespective of the source nodes characteristics for short periods of time. Therefore, these actions may be transient and not reflect a category membership defined by certain actions that will persist. These category labels can be related to consistent behaviors mapped to archetypes [21, 22] based on feature profiles which affect decisions. Therefore, this work also seeks to incorporate the fundamental formulations to the Katz centrality for information exchange and whether it can be used to assist the modeling of the overall feature influence propagation that determines a node's category label assignment (discussed in "Methodology" section).

Models for opinions dynamics find applications in marketing, politics, and urban changes amongst others. The work of [23, 24] models how exposure to new ideas can change the political opinions within a population but if left alone would revert to a preset independent position. This allows members with extreme opinions to sway the opinions of others iteratively in a simulation without the assumption that there is an actual alteration on internalized ideas. Along this principle, this paper explores how users with feature sets can have their categorical label identities changed without their own features requiring modification due to the exposure. This applies to situations where a node with a high proportion of links to nodes of a different label can adopt a different label as a result of the exposures. An alternative question from the opposite perspective is whether a node taking the role of an influencer, with dominating feature levels, can influence a set of different nodes through direct or indirect paths resulting in them taking upon a different label.

The methodology of the Simple Graph Convolutional Neural Network (SGC) [25] (described in more detail in "Methodology") presents an intuitive, simple, and expressive formulation for learning these latent representations of the nodes labels, which builds upon the general theory of graph convolutional networks [26]. What makes this methodology appealing is that the operations are linear between the adjacency matrix, the features, and the parameters prior to the use of the softmax function. This makes it an ideal candidate to work with in exploring different applications of its formulation as the feature projections are linear and that the adjacency matrix is clearly an operation aggregating feature information of the vicinity of nodes (number of 'hops'). The SGC is based on the Graph Convolutional Neural Network (GCN) [27] which fits a multilayered neural network to the features in a semi-supervised manner. A different approach taken by DeepWalk [28] is where truncated Markov chains are used to simulate exchanges in a radius of influence where the latent label distributions are produced based on the statistics of the chain state visits. Although the work of [27] does explore how the accuracy of the method with changes in the number of hidden layers employed, the authors do not question whether the ideal number of layers, corresponding to the number of k 'hops' (as stated in the paper), reflects the number of edge traversals used by the nodes

in establishing their label identity. If this is so, then the methodology can be employed to simulate label allocations for different numbers of edge traversals that can be expected to occur over time.

Diffusion simulation models for predicting changes typically involve simulations based on algorithms that capture information exchange over individual nodes and require a large number of iterations representative of social interactions where the mean statistics enable a label distribution to be obtained, [29, 30]. The approach taken here utilizing the SGC formulation (described in "Methodology" section) allows a different nature of a simulation of node label states based on the number of walks taken between nodes which influence each other. The results (shown in the Results section) demonstrate how labels can change based on the size of the influence neighborhood, and result in oscillations of assignments before convergence is obtained by a common label; as is often the objective in stochastic simulations of influence based on message propagation. The conclusions (in the Conclusion section) discuss the merits of this approach in light of the presented results and the potential applications. This provides a novel exploration of the SGC in terms of its ability to demonstrate the effect of the vicinity of influence for which a node is exposed to in terms of the label classification. This can be useful for answering questions regarding which users once included in a field of influence will alter a node's category label. The Results will present the exploration of a set of simulations on real and synthetic data where the label change is apparent given changes in the number of hops messages are permitted to travel as a change in the vicinity.

Methodology

A graph is defined as $\mathcal{G} = (\mathcal{V}, \mathbf{A})$ where $\mathcal{V}$ is the vertex set for the nodes $v_i \in \{v_1, \ldots, v_2\}$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$ being the adjacency matrix representing the node interconnectivity. Each entry in $\mathbf{A}$ is denoted as a_{ij} and can take different values for the weight between each node (v_i and v_j) and the weight for missing edges is 0, $a_{ij} = 0$. The degree matrix of the adjacency matrix is $\mathbf{D} = diag(d_1, \ldots, d_n)$, where $d_i = \sum_j a_{ij}$, which is the sum of the elements along the rows of the adjacency (outward edge summation for each node). Each node has a feature vector to represent its characteristic attributes and is identified as $\mathbf{x}_i \in \mathbb{R}^d$ and the set of the feature vectors stacked in a matrix is $\mathbf{X} \in R^{n \times d}$. In $\mathbf{X}$, there are n rows where each row is a feature vector belonging to a specific node and each column indexes a particular feature. The label set for which a node can be classified as it is unknown, and the ground truth has it as a 1-hot encoded vector $y_i \in \{0, 1\}^C$ (where C is the number of labels creating a C-dimensional vector). The node feature vectors too can contain 1-hot encoded feature variables where they are categorical. As will be seen, the increase in the number of the columns to facilitate this does not prohibit the methodology from operating consistently with the feature representations.

In [25], a detailed description of how the features are *propagated* and averaged (weighted) in the local neighborhood of v_i is presented. Considering how CNNs [31] represent feature transformations through multiple layers, at layer 0, the input data are the feature projection $\mathbf{H}^{(0)} = \mathbf{X}$ without network information. Each layer k depends upon the previous layer $k - 1$, so that these hidden layers average the feature representations consecutively from the node neighborhoods, defined by the adjacency matrix, which multilayer perceptrons do not perform. The feature propagation is performed via:

$$\bar{\mathbf{h}}^k \leftarrow \frac{1}{d_i+1}\mathbf{h}_i^{(k-1)} + \sum_{j=1}^{N} \frac{a_{ij}}{\sqrt{(d_i+1)(d_j+1)}}\mathbf{h}_j^{(k-1)}. \tag{1}$$

The component from the diagonal matrix having an increment is understood when noting how self loops are added to the adjacency which assists later on in calculating the matrix powers entries with zero edges. It can be seen how for each individual node in the network at each layer, there is an averaging of the features from all the other nodes $\mathbf{h}_j^{(k-1)}$ in the previous layer and from its own previous values $\mathbf{h}_i^{(k-1)}$. Another detail is worthy to notice how the larger d_j is, the number of outgoing edges of v_j, there will be a reduction on its ability to influence v_i as a type of 'dilution' of influence over more nodes occurs. The normalized adjacency matrix $\mathbf{S}$ will be used, $\mathbf{S} = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}$ where $\tilde{\mathbf{A}}$ denotes the adjacency matrix when self loops are added, and $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ ($\tilde{\mathbf{D}}$ becomes the degree matrix for $\tilde{\mathbf{A}}$). This layer connection can be shown by the relation, $\bar{\mathbf{H}}^{(k)} \leftarrow \mathbf{S}\mathbf{H}^{(k-1)}$. Nodes with large weighted edges between themselves will have a greater spread of the averaged features, so that the label outcomes will tend to converge to a similar projection.

A weight matrix, $\Theta^{(k)} \in \mathbb{R}^{n\times C}$, provides the parameters for the elements of the features to be scaled as a linear projection for each node and this is then input to a nonlinear activation function (such as ReLU, sigmoid, or the softmax) to produce a class label distribution. Each column of $\Theta^{(k)}$ corresponds to a vector of weights for a particular class's projection. Feature vectors of nodes $\mathbf{x}_i$ belonging to a class c are expected to have the largest projection value for that column $\mathbf{y}_c$. The classifier of the SGC, or GCN with one layer (1-hop), has the predictions for node labels in a matrix $\hat{\mathbf{Y}} \in \mathbb{R}^{n\times C}$ and $\hat{y}_{ic}$ denotes the probability for a node i belonging to class $c \in C$ via:

$$\hat{\mathbf{Y}} = \text{softmax}\Big(\mathbf{S}\mathbf{H}^{(1)}\Theta^{(1)}\Big). \tag{2}$$

What can be seen is that the linear projections of each node's feature vector along each class is produced and then the local averaging of those projections for each node's neighborhood is found via $\mathbf{S}$ from which the probabilities for each node class membership are found. If taken for subsequent iterations, this performs as the GCN, $\hat{\mathbf{Y}} = \text{softmax}\big(\mathbf{S}\mathbf{H}^{(K-1)}\Theta^{(K)}\big)$ where now K acts as a parameter for the number of hidden layers (here, the exponents in parentheses represent layers rather than exponents). In this formulation, each subsequent hidden layer is averaging the class membership for the nodes after another projection iteration with a new weight matrix for that layer, and the same normalized adjacency matrix. A key understanding is that these layers each average neighbors 1-hop from each node (walks and paths of length 1), so that at K layers each node is receiving feature projection information from k-hops away. The SGC performs a linearization by removing the nonlinearity (use of 'softmax') between the layers, so that there is a series of regressions only prior to the use of softmax; $\hat{\mathbf{Y}} = \text{softmax}(\mathbf{S}(\ldots(\mathbf{S}(\mathbf{S}\mathbf{X}\Theta^{(1)})\Theta^{(2)})\ldots\Theta^{(K)}))$ (inner matching parenthesis show the linear layers). This can be simple using the associative property, so the weight matrix can be equivalently represented by another different matrix and the normalized adjacency matrix by raising it to the power of the number of hops:

$$\hat{\mathbf{Y}} = \text{softmax}\Big(\mathbf{S}^K \mathbf{X}\Theta\Big), \tag{3}$$

which produces the formulation of the SGC. Without the adjacency matrix, the formulation becomes logistic regression, $\hat{\mathbf{Y}} = \text{softmax}(\mathbf{X}\Theta)$, or with an adjacency matrix without any off diagonal entries (setting $k = 0$). This allows the methodology to run very quickly compared to other methods requiring optimization of multiple weight matrices. In this work, we explore the effect of K on the label allocations not in terms of the accuracy but in the nature of the extension to the neighborhood of influence:

$$[\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_K] = \Big[\text{softmax}\Big(\mathbf{S}^1\mathbf{X}\Theta\Big), \text{softmax}\Big(\mathbf{S}^2\mathbf{X}\Theta\Big), \ldots, \text{softmax}\Big(\mathbf{S}^K\mathbf{X}\Theta\Big)\Big]. \tag{4}$$

The authors of the SGC note that it addresses a pitfall found in the GCN where large numbers of layers (more than 3) induce reductions in the accuracy which the SGC avoids, allowing for a greater exploration of the influence due to the removed range restriction. The Results section demonstrates how this exploration can be used to see how influence changes labels over the neighborhood size with K.

From the formulation of Eq. 3, we notice that the inclusion of the component $\mathbf{S}^K$ differentiates the method from the established method of logistic regression (the Results section will provide a visual depiction of the effect of K on $\mathbf{S}$). Using the adjacency matrix raised to a power K, $\mathbf{A}^K$ produces a K step chain of edge traversals (permitting node revisits) called 'walks' [32]. The Katz centrality measure, [15], attaches a constant α representing the probability of an effective use of an edge (set to 0.8 as recommended from experiment in [17, 18]). If then a walk of length K takes place, there is a probability α^K of it being effective:

$$\alpha^1\mathbf{A}^1 + \alpha^2\mathbf{A}^2 + \cdots + \alpha^k\mathbf{A}^k + \cdots = (I - \alpha\mathbf{A})^{-1}. \tag{5}$$

The last term is the matrix resolvent which is not used in this work, but in the manner which DeepWalk takes truncated Markov chains here the approximation for a limited k is taken:

$$K_{katz} = \alpha^1\mathbf{A}^1 + \alpha^2\mathbf{A}^2 + \cdots + \alpha^K\mathbf{A}^K. \tag{6}$$

This is approximation which is also employed in commercial practice to avoid the large matrix inversion cost [18] and can produce analogous results. In the SGC, the K-hop chains lack such an explicit representation of a decay effect with length, but to some extent, it is present implicitly with the consecutive multiplications of a set of normalized numbers forcing the smaller ones to more quickly shrink. This relative value that eventually would become evenly dispersed across the columns for large K does show that the local proximity of nodes would have diminishing influence over a continuous long stretch of influential passes. Although this makes sense, it does not account for the 'first mover' advantage [33, 34], since local nodes would provide influence prior to more distant nodes. The formulation introduced for this nature of a simulation is:

$$\mathbf{Y}_k = \text{softmax}\left(\sum_{k=1}^{k} \alpha^k \mathbf{S}^k \mathbf{X}\Theta\right), \tag{7}$$

and we keep Θ fixed, although the formulation assumes the matrix for the specific 'k' to be used in the semi-supervised scheme, but we assume that the weights are constant for the principles of how those features combine irrespective of the data fits that may direct it to be otherwise. This equation provides an adaptation of the original SGC formulation for accounting of the different walk lengths (k-hops) that changes the vicinity of influence. Depending upon the penalization of the walk length and the maximum walk length permitted, it can be examined which nodes have a change in their inferred labels. In practice for threads that do not cascade, a low value of K may be relevant, and for longer threads of repropagated information, a larger K would be suitable. Previous research has not directly investigated the impact of K or provided a formulation for the combined impact of different values of K upon label inferences.

This takes inspiration from agent-based systems modeling where the simulations frequently use a constant set of dynamics for the agent behaviors from their environment [35]. This approach is also found in other sociological investigations such as with the Schelling model [36] where the dynamics for movements of people depends upon a prefixed set of parameterized decisions. Although data are not used to infer the parameter values of Θ, previous research [37] has looked at random weights to serve as feature extractors, as also noted in [27]. The experiments will also include $k = 0$ which results in the $\mathbf{S}$ matrix being the identity matrix, and then, the formulation produces the results of logistic regression, so the nodes use only the information of their own features.

Results

Figure 1 shows three of the four networks that the methodological approach will be applied to. Subfigure a) shows a synthetic network of an approximately block diagonal adjacency matrix. There is a set of three sparsely connected components with random perturbations of the densely connected components to create a non-disjoint graph. The purpose of this network is to investigate weather influence propagation can cross the few *boundary nodes* linking communities. This has applications in protecting communities online from hazardous or inappropriate content. Subfigure b) shows the network diagram of a real network adjacency matrix produced from the 'Zachary Karate Club' dataset, [38]. The importance of experiments is to see if the iterated SGC with changes in the values of k will preserve the segmentation of the influences. The iterated approach from Eq. 4, and the cumulative approach of Eq. 7 will be applied to these data. Subfigure a) shows a ring network where each node has 2 neighbors and the colors represent the value of the single feature variable that ranges between red and blue where grays correspond to values around zero. The range from blue to red is representative of a political leaning on the 2 party political spectrum and the application seeks to see how a consensus arises from larger values of K and if from lower values oscillations appear from local intermediate consensus outcomes.

Applying the iterated SGC on a ring network

The examination here looks at the context where a set of nodes are placed in a ring network (Fig. 1. Each node has a single feature variable value in $[-1, +1]$. This application can potentially correspond to political voting opinions on the spectrum typically encountered between 'liberal' and 'conservative' where the values place the node's

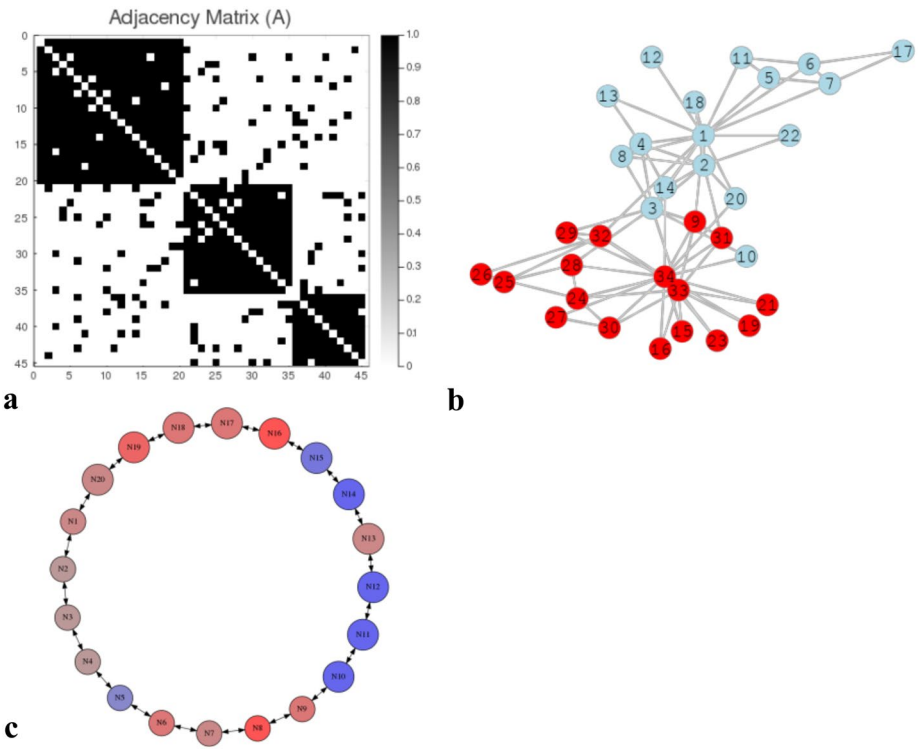

Fig. 1 Different networks used in exploring the methodology. **a** A network set of three sparsely connected clusters with random perturbations as an adjacency matrix. **b** The layout of the Zachary Karate Club and the class membership labels allocated from applying community detection. **c** A ring network that will be used, and the colors depict the value of the single feature variable each node contains (from red to blue with grays for in between values)

leaning or possibly neutral position. The methodology of Eq. 4 is applied to see how the influence of nodes in a changing vicinity can alter the decisions. The values are continuous, but since decision of voting take a discrete value, the rounded values will also be explored. The network has 20 nodes and 20 edges, and the density is 0.105 (3 significant figures).

Figure 2 shows the results of the ring network examination. Subfigure a) shows the group membership from the rounding to the nearest $[-1, +1]$ and the proportions for each group stacked per value of k. It can be seen how with increasing vicinity, the dominant group influences the full network till consensus producing a single outcome across all nodes. Subfigure b) shows how the values change with k and how the changes are increasingly smooth amongst the locality of the nodes with larger k values. The values per k are not rounded to $[-1, +1]$ and the initial values at $k = 0$ can be seen as well as an approximately uniform final distribution. Subfigure c) shows the same as b), but the values are rounded to either -1 or $+1$. It can be seen how sporadic changes can appear from local associations at low k values. For Subfigure d), the changes in the group memberships for the opinions represented in each group are shown, so that the exchanges can be tracked. The changes across the subplot cancel each other out as node assignments are exchanged between groups. Variability is

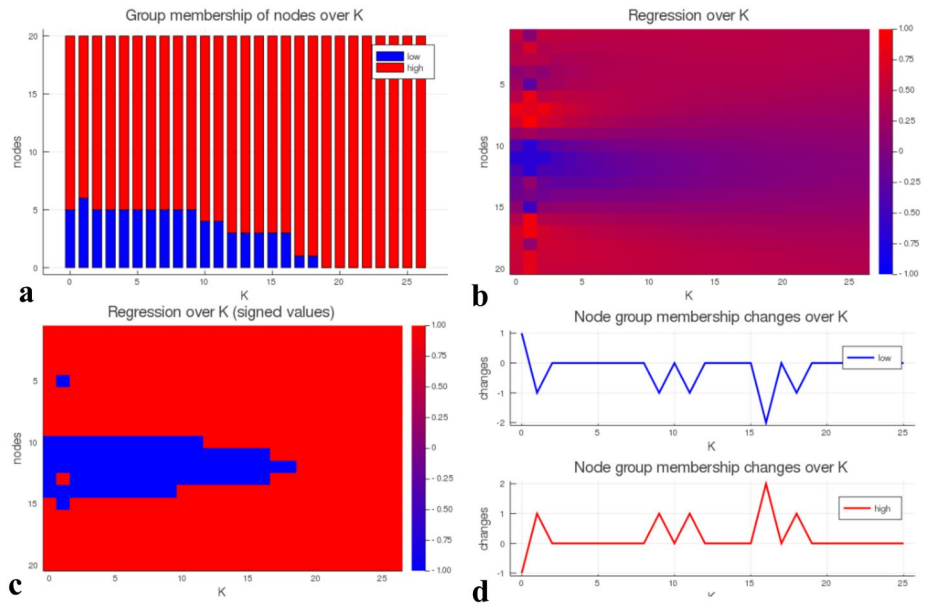

Fig. 2 Application of Eq. 4 to a ring network. **a** The proportions of the group numbers over the values of *k*. It can be seen how with increased *k*, one group dominates producing consensus. **b** The values of the linear projection for each node as they change with *k*. **c** The value rounded to the nearest $[-1, +1]$ to see what 'decision' would be made by each nodes due to the influence vicinity changes. **d** Has 3 subplots for the counts of the changes of the node labels over *k* with the vicinity changes

presented just as it would be expected for a stochastic simulation as local network arrangements would arrive at local agreements that differ from the macroscopic ones.

Application to a synthetic dataset with three connected clusters of nodes

In this exploration, the network defined by the adjacency matrix shown in Subfigure a) of Fig. 1 is used. There are three clusters of nodes differences in the number of nodes in each cluster. There is a dense edge set between nodes of the same labels, and perturbations are applied randomly to connect all the components together. The network has 45 nodes and 391 edges, and the density is 0.395 (three significant figures). The weight matrix, Θ for the three class feature projections is:

$$\begin{bmatrix} 0.0 & 0.0 & 1.0 \\ 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \\ 0.0 & 1.0 & -1.0 \\ 1.0 & 0.0 & -1.0 \\ 1.0 & 1.0 & -1.0 \end{bmatrix}. \tag{8}$$

For the feature matrix, $\mathbf{X}$, the first three rows correspond to a categorical variable with 1-hot encoding, the rows 4–6 to a discrete choice of numbers in the set $[-1, 0, 1]$, and the last value is a random normal addition to the mean value of that feature.

Figure 3 shows the results for running the indexed SGC proposed in Eq. 4. Subfigure a) shows the group membership proportions over *k*. What can be seen in is that the smallest cluster is dominated earlier on by the largest cluster losing its identity and the

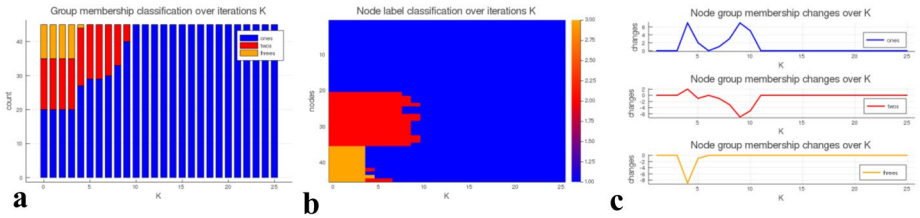

Fig. 3 Application to a network with three dense clusters of nodes. **a** The group membership of the nodes over k and how the vicinity allows the influence of the largest group to dominate the network labels of all other nodes. **b** The changes to each individual node and how the red group can find a k values to exploit a local network structure where it briefly dominates some orange nodes more than the more numerous blue group. **c** The individual change of the group over k

second lowest cluster size required a larger vicinity value of k for the largest cluster to overwhelm its local influence. Effectively, the larger k results in local feature projection becoming less than the dominant group. This can occur, even though the cross community connections which are relatively sparse, so that different groups can find their group membership altered from information exchange based on these edges also referred to as 'boundary nodes' [39–41]. Subfigure b) shows the changes in the node membership which can be tracked over k per node. It can be seen how the red group can exploit local segments of the network to override the influence of the dominant group and the smallest group (orange). Subfigure c) shows the changes over the three groups per value of k and the increments for the group sizes.

Exploration of the indexed SGC with application to the Zachary Karate Club dataset

Here, we explore how the model of influence exchanges over k takes place when the iterations are governed by Eq. 4 which is the indexed trace of the SGC over k. The network used is the Zachary Karate Club network with the network adjacency used from Subfigure b) in Fig. 1. There is no accumulative effect over the changes of k as each iteration is computed independently and is useful to see the state of the influence spread at each value k. What is examined is if a central node (node 1) and a peripheral node (node 27) taking influential values (large in magnitude) can affect a large part of the network over k. To start the network in such a state the class membership is modified, so that node '1' and '27' belongs to a group 3, and the feature vector is sampled from a different generator. Group 1 samples are taken from $[1+\mathcal{U}(0,10), 1+\mathcal{U}(0,4), \mathcal{U}(0,1)]$, Group 2 samples are taken from $[1+\mathcal{U}(0,4), 1+\mathcal{U}(0,10), \mathcal{U}(0,1)]$, and for Group 3 $[\mathcal{U}(0,1), \mathcal{U}(0,1), 30+\mathcal{U}(0,60)]$. The reason Group 3 has been allocated a large expected value is that the features that are characteristic for it are meant to be more influential, and that this is a manner in which the strength of a node's influence can be intuitively represented and produce an effect in this formulation. Situations for this can be charismatic characters, expert opinion holders, or those with dominating personalities. Therefore, this phenomenon has an intuitive representation by scaling certain feature variable values. The weight matrix, Θ, is set to be the identity matrix. The network has 34 nodes and 78 edges, and the density is 0.139 (3 significant figures).

Figure 4 shows the results of using the formulation of the SGC in Eq. 4 where the x-axis is k and shows how this parameter changes the number of 'hops' for which the

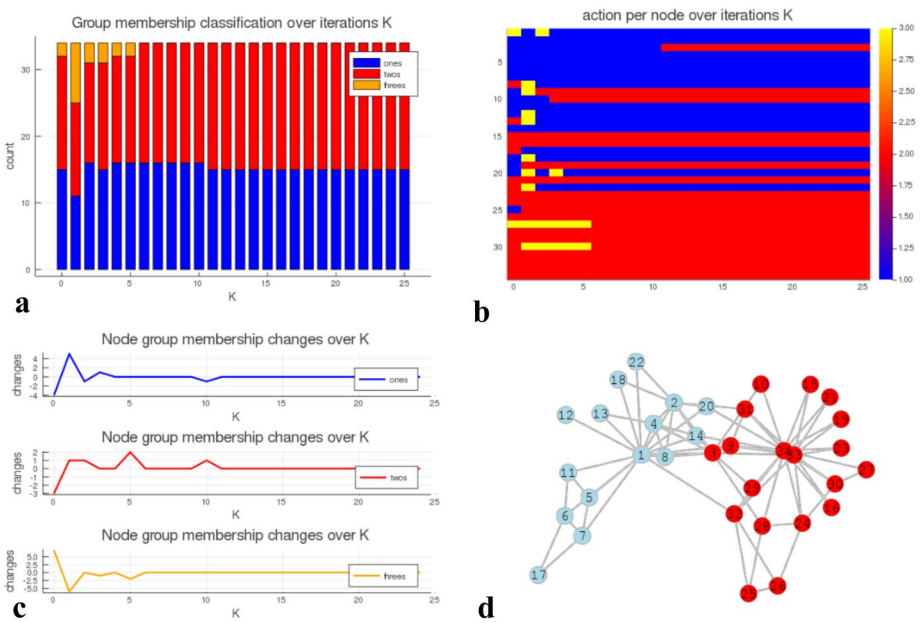

Fig. 4 Two nodes represent a third group member in the Zachary Karate Club dataset. Here, we can see the simulation (using the Zachary network) and how an attempt by a single influential node to spread its influence as the neighborhood of interaction increases. **a** The proportion of nodes per group over k and **b** the specific node changes. **c** The number of the changes to the group memberships over k. **d** The layout of the Zachary Karate Club and the class membership labels allocated from the simulation at the final step

walks produced that is akin to a simulation based on the vicinity changes of node influence. Subfigure a) shows the proportions of the nodes which belong to each of the three groups for a value of k. Group 3 ('orange') starts off as a comparatively very small subset of the node population, and then, the influence grows with k for a few steps. It can be seen how the two different groups remain stable regardless of the vicinity increases which provides insight that this network which arose organically is robust and, therefore, would maintain polarity and eventually separate. The experiments with the three clusters did not manage to ensure this, even though the boundary nodes (cross community bridges) were few. This provides insight for how to maintain social integrity from polarizing individuals [42]. Subfigure b) presents the same information but the y values correspond to node positions, so the changes in node classification can be more precisely tracked for the simulation trace. It can be seen how the group 3 nodes can sporadically make local changes and that the red and blue maintain relative stability. Subfigure c) shows the number of node changes over k as the index of the simulation where each subplot shows the difference in the number of nodes for each group. The summation over the separate lines cancel each other out. Subfigure d) shows the network group membership in a network diagram at the final k value.

Exploring the cumulative iteration of influence exchanges on the Zachary Karate Club dataset

Here, we explore how the model of influence exchange over k takes place when the iterations are governed by Eq. 7 and how it compares to the results obtained in the

previous subsection which does not use accumulated local influences in the iteration or a decay in the influence ability with distance. The is set to $\alpha = 0.8$ (as prescribed in [18]) and the network used is the Zachary Karate Club network with the network adjacency shown in the network diagram from Subfigure b) in Fig. 1. The network class membership is modified, so that node '1' and '27' belongs to group 3, so that the feature vector it is sampled from has a different generator. Group 1 samples are taken from $[1+\mathcal{U}(0,10), 1+\mathcal{U}(0,4), \mathcal{U}(0,1)]$, Group 2 samples are taken from $[1+\mathcal{U}(0,4), 1+\mathcal{U}(0,10), \mathcal{U}(0,1)]$, and for Group 3 $[\mathcal{U}(0,1), \mathcal{U}(0,1), 30+\mathcal{U}(0,60)]$. The reason Group 3 has been allocated a large expected value is that the features that are characteristic for it are meant to be more influential and that this is a manner in which the strength of a node's influence can be intuitively represented. Situations for this can be charismatic characters, expert opinion holders, or those with dominating personalities. The weight matrix, Θ, is set to be the identity matrix. The impact of the formulation of Eq. 7 in comparison to Eq. 4 can be seen where the walks of different lengths accumulate to produce an overall influence for each node's classification and the walk lengths are penalized with larger k.

Figure 5 shows the results of using Eq. 7. Subfigure a) shows the proportions of the nodes which belong to each of the three groups. It can be seen how Group 3 begins as a comparatively very small subset, and then, the influence continues to grow with k in contrast when the non-accumulative approach is taken. The increase in Group 3 and the corresponding decrease in Group 1 show that the change of label membership is restricted to one side of the network and that the neighborhood vicinity is too far and weak for the influence to extend beyond it. The non-central node 27 within the group 2 vicinity does not manage to divert the label classifications of many of its surrounding nodes in the same way node 1 does. This provides insight for how to maintain social integrity from polarizing individuals [42]. Subfigure b) presents the same information, but the y values correspond to node positions, so the changes can be more precisely tracked for the simulation trace. It can be seen how there are sporadic changes, but a trend is maintained which is not seen the results of the previous subsection and is attributed to the decay of influence for large k values by the factor of α^k. Subfigure c) shows the number of node changes over k as the index of the simulation where each subplot shows the difference in the number of nodes for each group. The summation over the separate lines cancel each other out. Subfigure d) shows the network group membership at the final value of K. A key difference is that the lower values of k allow nodes to adapt more locally when node 1 has less of the influence of the opposing group's central nodes.

Exploring the influence behaviors on the Eu-Email-core dataset

The approaches defined in Eq. 4 and 7 are applied to the dataset of 'email-Eu-core' [43, 44] which is part of the SNAP collection [45]. These network data are produced from email exchanges between large European research institutions. There are 1005 nodes and 25571 edges, and a density of 0.0253 (three significant figures). Along with this network is a set of labels for the departments of which there are 42. The first 10 departments are taken to be part of Group 2 ('red' category with 431 nodes), the departments 11 & 12 to Group 3 ('orange' category 32 nodes), and the rest of the department nodes belong to Group 1 ('blue' category). Group 1 features

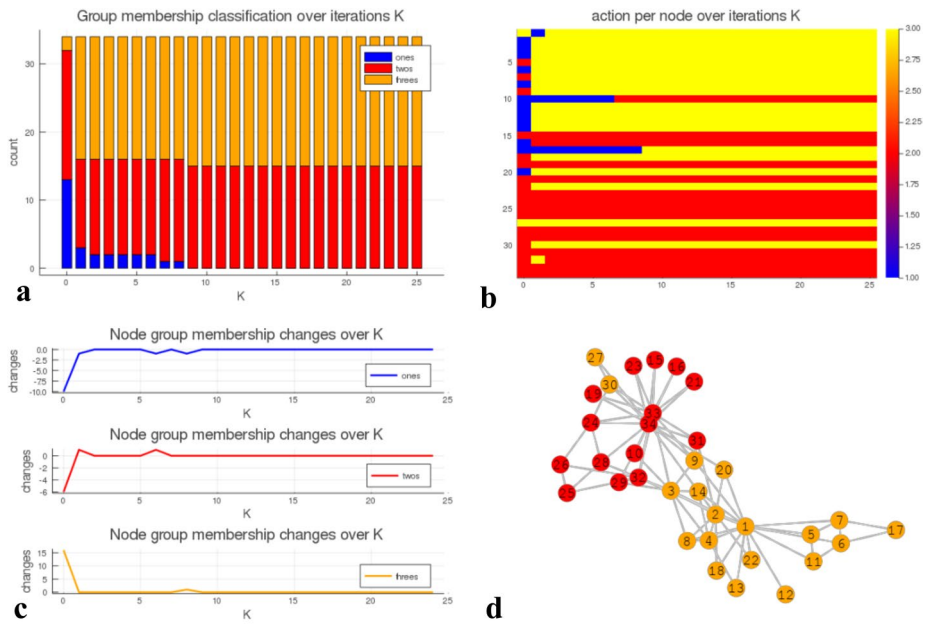

Fig. 5 A single very influential node and a peripheral node represent a third group member in the Zachary Karate Club network to see if their influence can dominate nodes. Here, we can see the cumulative iteration applied of Eq. 7 (using the Zachary network) and how node '1' and '27' can spread their influence as the neighborhood of interaction increases. **a** The proportion of nodes per group and how one group identity ceases to exist, but the third group can grow and become stable. **b** The specific node changes along *k*. **c** The number of the changes to the group memberships over *k*. **d** The network diagram of the Zachary Karate Club, and the class membership labels allocated from the simulation at the final step

are taken from $[1+\mathcal{U}(0,10), 1+\mathcal{U}(0,4), \mathcal{U}(0,1)]$, Group 2 features are taken from $[1+\mathcal{U}(0,4), 1+\mathcal{U}(0,10), \mathcal{U}(0,1)]$, and for Group 3 $[\mathcal{U}(0,1), \mathcal{U}(0,1), 100+\mathcal{U}(0,100)]$, so that Group 3 has the strongest influence score value, but are the least numerous of the groups. The simulations use the proposed modification of the SGC to see how influence can be spread using the number of k-hops as 'walks'. This is also compared to the simulations where the nodes for each group are chosen based on degree centrality, so that the first 32 nodes belong to Group 3, the following 431 nodes to Group 2, and the rest of the nodes to Group 1.

Figure 6 presents the results of using this dataset. In subfigures a) and b), the department separations are used, and in c) and d), the degree centrality rank is used to allocate nodes to each group. Subfigures a) and c) show the results of using Eq. 4, and in b) and d) using Eq. 7. From a) and b), it can be seen how the influence of the orange group can extend to a larger group and to a lesser extent the red group for a few hops, and then, there is a decay as the k continues to increase as most nodes of the network are then incorporated in each node group classification. The effect of the decay and accumulation of Eq. 7 allows the walks of larger lengths to have a reduced influence and the effect of the shorter walks to have a carry-on effect. This produces the effect that there is greater stability in the influence of the classifications over the value of k. As an application, it demonstrates how a relatively small department can have a large impact upon an institutional network of academics. The results of c) and d) mirror this effect, but since there is also a topological positioning for the influential groups to spread their influence, the changes are larger. The most influential nodes could be expected to be part of the nodes

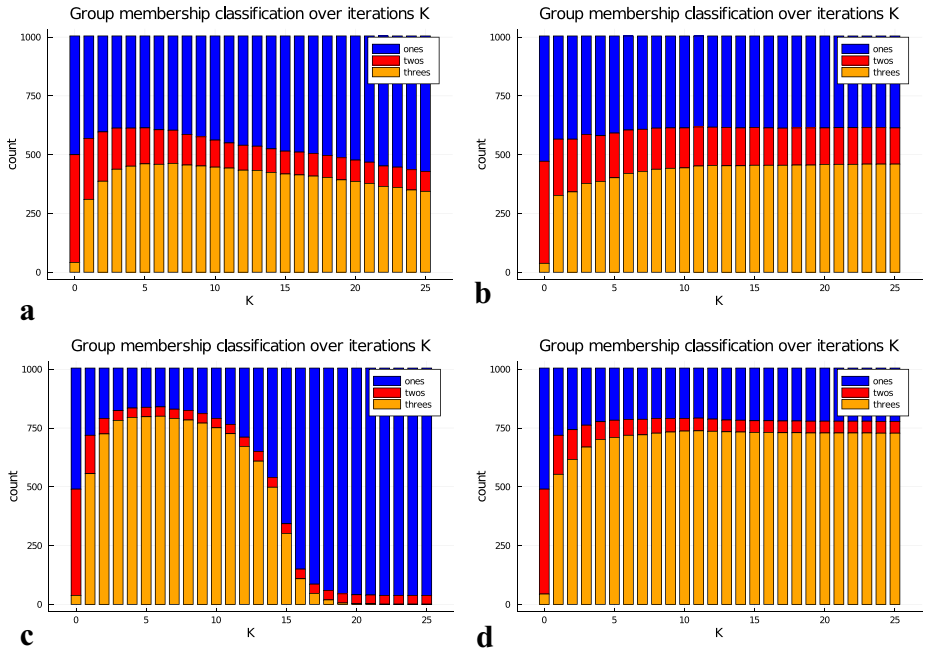

Fig. 6 This figures presents the results of applying the approaches to the dataset of 'email-Eu-core' [43, 44] part of the SNAP collection [45]. In **a**, **b** the nodes in the first ten departments belong to category 'two' (red), the nodes in departments 11 & 12 to category 'three' (orange), and the rest are in category 'one' (blue). There is asymmetric influence with category 'three' having the most which affects the label association with different values of *k*. For **a**, Eq. 4 is used, and in **b**, Eq. 7 is used. With **c**, **d**, the same number of nodes for the different categories is used, but instead of choosing those nodes by department, they are selected from the ordered degree centrality rank. Both pairs of plots demonstrate how Eq. 4 depicts a mechanism for influence spread upon the number of walks, and with Eq. 7 the decay over the length with the carry-on effect

with the largest number of direct connections and even if they have the same number as for a) and b), we see a greater number of nodes are influenced for lower values of k in c), since the orange group has more direct edges to nodes to influence them. As before, in d), the accumulation allows for the effect of the shorter walks to take a greater role in the category assignment in the presence of longer walks with large k. This further supports that approach to using the modified SGC can simulate the spread of influence and for groups with different feature values depending upon the walk length.

Figure 7 shows the changes of the results based on the value of α in Eq. 7 which are now set to take on the values of [0.25, 0.5, 0.75, 1.0]. The figure shows the range tested on Fig. 6d) and it can be seen how with the exception of $\alpha = 1.0$. the increases allow for the influence of group three to extend further. This is due to the large feature value characteristic of that group having a smaller decay further away. For the larger values of k, the blue group ('one') increases its dominance with long-range non-penalized walks benefit its large group population size.

Conclusion

This work shows how the Simple Graph Convolutional Neural Network (SGC) of [25] can be used to explore how nodes influence each others' label allocation with changes in the neighborhood vicinity size. The parameter K allows the exploration to examine the

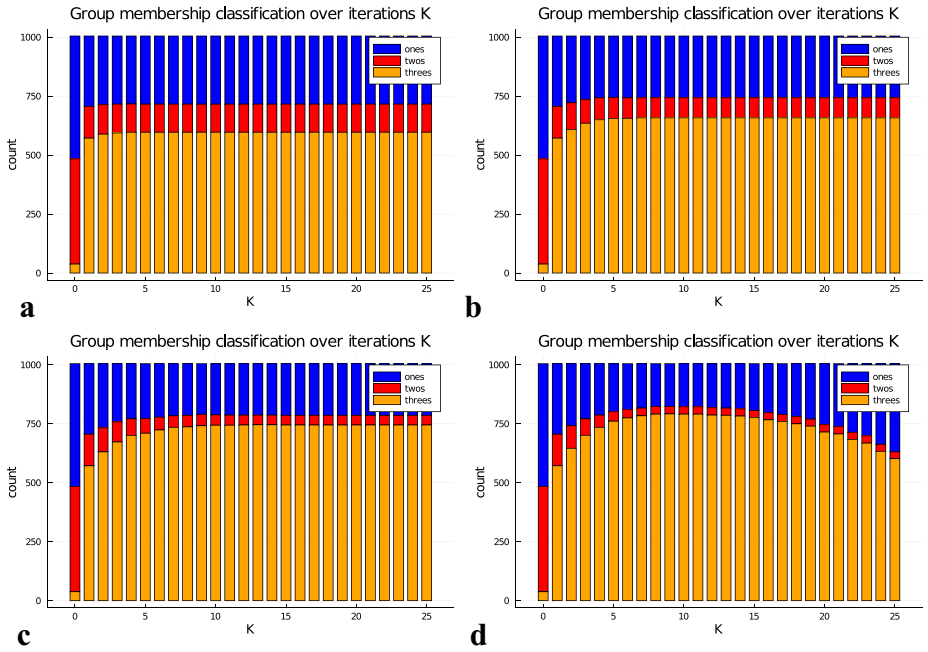

Fig. 7 This figure presents the results of examining the effect of the α parameter used in Eq. 7 to the dataset of 'email-Eu-core' [43, 44] part of the SNAP collection [45]. The different examinations **a**–**d** use the α values of [0.25, 0.5, 0.75, 1.0]. The scenario of Fig. 6d) is used. An increase in the value allows the influence of the orange group to extent further until for very large values of k where the blue group nodes influence for large k gain influence without the long walk decay and their numerical advantage can be applied

label allocations for a range of cases including what allocations would exist if nodes did not receive external influence from the network ($K = 0$ which causes the methodology to become equivalent to logistic regression). Large values would result in the local influence of a node to become diluted as communication from further parts of the network take a more equal role. The experiments show how local group memberships decrease and can cease to exist with large k values even when the number of edges allowing for those cross group connections is sparse. This adds value to the relevant research in boundary nodes (spanner nodes) [39–41] and how they can facilitate important exchange as a few edges can provide an ability for a larger group to dominate. A formulation proposed here allows the methodology to incorporate the important concepts that nodes further away have the walks between nodes penalized in proportion to that walk length, and that nodes from a closer vicinity will have a greater influence on other nodes due to nodes further away having delayed information arrivals.

Using this approach allows for an efficient mechanism for influence exchanges in a network with an intuitive and explainable set of terms in the formulation. A key feature is to examine the changes between labels over the influence vicinity to see if certain groups are vulnerable to become dominated by other labels in the connected cluster sets. The figures of the results show how changes as differences in the number of nodes included can be monitored over a k. Other use cases can see which consensus class labels are produced with large values of k, and to see if certain network configurations do provide an effective barrier against different group influences.

Future work could proceed by examining how representations can be made to handle known variable interactions in the feature matrix and how to perform dimensionality reduction on the feature set needed from the nodes.

Abbreviations
SGC: Simple Graph Convolutional Neural Network; GCN: Graph Convolutional Neural Network; CNN: Convolutional Neural Networks.

Acknowledgements
Not applicable

Authors' contributions
Conceptualization, AVM, DC, and KR; formal analysis, AVM; investigation, AVM, DC, and KR; methodology, AVM; software, AVM; supervision, AVM; validation, AVM, DC, and KR; visualization, AVM; writing—original draft, AVM; writing—review & editing, AVM.

Funding
This work was partially supported by grant FA8650-18-C-7823 from the Defense Advanced Research Projects Agency (DARPA). The views and opinions contained in this article are the authors and should not be construed as official or as reflecting the views of the University of Central Florida, DARPA, or the U.S. Department of Defense.

Availability of data and materials
The code used will be provided as a repository from GitHub at [46] with the project name, sgcCommunity. The implementation was tested with Julia 1.3.1 within a Jupyter Notebook running on Linux, and is offered with the MIT License open source license. Another requirement is the use of GraphViz [47].

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Department of Statistics and Data Science, University of Central Florida (UCF), 4000 Central Florida Blvd, Orlando 32816, USA. [2] Department of Computer Science, University of Central Florida (UCF), 4000 Central Florida Blvd, Orlando 32816, USA.

Received: 26 May 2020 Accepted: 4 March 2021
Published online: 17 March 2021

References

1. Newman M. Networks. Oxford: Oxford University Press; 2018.
2. Estrada E. The structure of complex networks: theory and applications. Oxford: Oxford University Press; 2012.
3. Euler, L. Solutio problematis ad geometriam situs pertinentis. Commentarii academiae scientiarum Petropolitanae, 1741;128–140.
4. Matai R, Singh SP, Mittal ML. Traveling salesman problem: an overview of applications, formulations, and solution approaches. Travel Salesm Prob Theory Appl. 2010;1:12.
5. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-L. The large-scale organization of metabolic networks. Nature. 2000;407(6804):651–4.
6. Roopnarine P. Graphs, networks, extinction and paleocommunity food webs. Nat Prec. 2010;15:1–1.
7. Dunne JA, Williams RJ, Martinez ND. Food-web structure and network theory: the role of connectance and size. Proc Natl Acad Sci. 2002;99(20):12917–22.
8. Smith EB, Brands RA, Brashears ME, Kleinbaum AM. Social networks and cognition. Annu Rev Sociol. 2020;46:87.
9. Schafer JB, Frankowski D, Herlocker J, Sen S. Collaborative filtering recommender systems. In: The adaptive web. Berlin: Springer, 2007, pp. 291–324.
10. Hamm JV. Do birds of a feather flock together? the variable bases for african american, asian american, and european american adolescents' selection of similar friends. Dev Psychol. 2000;36(2):209.
11. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Ann Rev Sociol. 2001;27(1):415–44.
12. Kossinets G, Watts DJ. Origins of homophily in an evolving social network. Am J Sociol. 2009;115(2):405–50.
13. Mantzaris AV, Higham DJ. Infering and calibrating triadic closure in a dynamic network. In: Temporal networks. Berlin: Springer, 2013, pp. 265–282.
14. Lobel I, Sadler E. Information diffusion in networks through social learning. Theor Econ. 2015;10(3):807–51.
15. Katz L. A new status index derived from sociometric analysis. Psychometrika. 1953;18(1):39–43.
16. Borgatti SP, Everett MG. A graph-theoretic perspective on centrality. Social Netw. 2006;28(4):466–84.
17. Grindrod P, Parsons MC, Higham DJ, Estrada E. Communicability across evolving networks. Phys Rev E. 2011;83(4):046120.
18. Laflin P, Mantzaris AV, Ainley F, Otley A, Grindrod P, Higham DJ. Discovering and validating influence in a dynamic online social network. Social Netw Anal Mining. 2013;3(4):1311–23.

19. Taecharungroj V. Starbucks' marketing communications strategy on twitter. J Mark Commun. 2017;23(6):552–71.
20. Zhang L, Zhao J, Xu K. Who creates trends in online social media: the crowd or opinion leaders? J Comput Mediat Commun. 2016;21(1):1–16.
21. Mirbabaie M, Bunker D, Stieglitz S, Deubel A. Who sets the tone? determining the impact of convergence behaviour archetypes in social media crisis communication. Inf Syst Front. 2019;14:1–13.
22. Narang K, Chung A, Sundaram H, Chaturvedi S. Discovering archetypes to interpret evolution of individual behavior. arXiv preprint arXiv:1902.05567 2019.
23. Cobb L. Mathematical Frontiers of the social and policy sciences. New York: Routledge; 2019.
24. Higham DJ, Mantzaris AV. A network model for polarization of political opinion? a3b2 show [editpick]? Chaos: an Interdisciplinary. J Nonlin Sci. 2020;30(4):043109.
25. Wu F, Zhang T, Souza Jr AHd, Fifty C, Yu T, Weinberger KQ. Simplifying graph convolutional networks. arXiv preprint arXiv:1902.07153 2019.
26. Zhang S, Tong H, Xu J, Maciejewski R. Graph convolutional networks: a comprehensive review. Comput Social Netw. 2019;6(1):11.
27. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv :1609.02907 2016.
28. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, 2014; pp. 701–710
29. Etzion D. Diffusion as classification. Organiz Sci. 2014;25(2):420–37.
30. Hajibagheri A, Hamzeh A, Sukthankar G. Modeling information diffusion and community membership using stochastic optimization. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013; p. 175–182
31. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86(11):2278–324.
32. Duncan A. Powers of the adjacency matrix and the walk matrix. 2004.
33. Kerin RA, Varadarajan PR, Peterson RA. First-mover advantage: a synthesis, conceptual framework, and research propositions. J Mark. 1992;56(4):33–52.
34. Varadarajan R, Yadav MS, Shankar V. First-mover advantage in an internet-enabled market environment: conceptual framework and propositions. J Acad Mark Sci. 2008;36(3):293–308.
35. Epstein JM. Agent_Zero: toward neurocognitive foundations for generative social science, vol. 25. Princeton: Princeton University Press; 2014.
36. Gracia-Lázaro C, Lafuerza LF, Floría LM, Moreno Y. Residential segregation and cultural dissemination: an axelrod-schelling model. Phys Rev E. 2009;80(4):046123.
37. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010; pp. 249–256
38. Zachary WW. An information flow model for conflict and fission in small groups. J Anthropol Res. 1977;33(4):452–73.
39. Mantzaris AV. Uncovering nodes that spread information between communities in social networks. EPJ Data Sci. 2014;3(1):26.
40. Girvan M, Newman ME. Community structure in social and biological networks. Proc Natl Acad Sci. 2002;99(12):7821–6.
41. Long JC, Cunningham FC, Braithwaite J. Bridges, brokers and boundary spanners in collaborative networks: a systematic review. BMC Health Services Res. 2013;13(1):158.
42. Garibay I, Mantzaris AV, Rajabi A, Taylor CE. Polarization in social media assists influencers to become more influential: analysis and two inoculation strategies. Sci Rep. 2019;9(1):1–9.
43. Leskovec J, Kleinberg J, Faloutsos C. Graph evolution: densification and shrinking diameters. ACM Trans Knowl Disc Data (TKDD). 2007;1(1):2.
44. Yin H, Benson AR, Leskovec J, Gleich DF. Local higher-order graph clustering. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, 2017; pp. 555–564
45. Leskovec J, Krevl A. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data 2014.
46. Mantzaris AV. SGC community influence. GitHub 2020
47. Ellson J, Gansner E, Koutsofios L, North SC, Woodhull G. Graphviz—open source graph drawing tools. In: International symposium on graph drawing. Berlin: Springer; 2001, p. 483

Publisher's Note