machine learning & knowledge extraction

MDPI

Article

Exploiting Weak Ties in Incomplete Network Datasets Using Simplified Graph Convolutional Neural Networks

Neda H. Bidoki [1], Alexander V. Mantzaris [2] and Gita Sukthankar [1,*]

[1] Department of Computer Science, University of Central Florida (UCF), Orlando, FL 32816, USA; nedahaji@cs.ucf.edu

[2] Department of Statistics and Data Science, University of Central Florida (UCF), Orlando, FL 32816, USA; alexander.mantzaris@ucf.edu

* Correspondence: gitars@eecs.ucf.edu

Received: 20 April 2020; Accepted: 20 May 2020; Published: 21 May 2020

Abstract: This paper explores the value of *weak-ties* in classifying academic literature with the use of graph convolutional neural networks. Our experiments look at the results of treating *weak-ties* as if they were *strong-ties* to determine if that assumption improves performance. This is done by applying the methodological framework of the Simplified Graph Convolutional Neural Network (SGC) to two academic publication datasets: Cora and Citeseer. The performance of SGC is compared to the original Graph Convolutional Network (GCN) framework. We also examine how node removal affects prediction accuracy by selecting nodes according to different centrality measures. These experiments provide insight for which nodes are most important for the performance of SGC. When removal is based on a more localized selection of nodes, augmenting the network with both *strong-ties* and *weak-ties* provides a benefit, indicating that SGC successfully leverages local information of network nodes.

Keywords: graph convolutional neural networks; weak ties; social networks; collective classification

1. Introduction

In addition to providing entertainment and social engagement, social networks also serve the important function of rapidly disseminating scientific information to the research community. Social media platforms such as ResearchGate and Academia.edu help authors rapidly find related work and supplement standard library searches. Twitter not only serves as an important purveyor of standard news [1] but also disseminates specialty news in fields such as neuroradiology [2]. Venerable academic societies such as the Royal Society (@royalsociety) now have official Twitter accounts. Shuai et al. [3] discuss the role of Twitter mentions within the scientific community and the citations that create a topological arrangement between scientific publications. Given that scientific articles possess the potential to change the landscape of technology, it is important to understand the information transference properties of academic networks: can techniques originally developed for social networks yield insights about scientific networks as well?

Pagerank [4] produced a revolution in the ability to search through the myriad of webpages by examining the network structure for relevance. This concept has been applied to citation networks in academic literature, which conceptually has many overlaps with the interlinking of websites, and Ding et al. [5] apply the Pagerank algorithm to citations. The same first author extends this work to investigate endorsement in the network as well [6]. Endorsement as a process produces an effect that not only allows readers to navigate essential information, relevant overlaps or apply appropriate

credit but also amplifies readership which is a dynamic seen frequently in online social processes. These links are direct links and also referred to as *strong-ties* [7].

Indirect or *weak-ties* have also been identified as important in social networks, most notably in Granovetter's seminal work: "The strength of weak ties" [8]. These indirect edges can be created through edges that span different communities (clusters of nodes) acting as 'bridges'. They can also be the result of 'triangulation' where 'friends-of-friends' produce a link due to the common connection they share. Figure 1 shows an example of a *weak-ties* connection between nodes A-C with a dotted edge representing connectivity due to a shared connection with node B. It can be said that A-C are connected from friends-of-friends, or triangulation creating a *weak-tie*. The work of [9] applies this concept to predicting edge production in a social network of professional profiles and shows that the explicit modeling of this triangulation dynamic leads to improved performance.

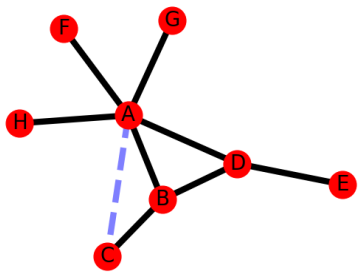

Figure 1. This figure illustrates how *weak-ties* can be produced through triangulation.

For our study on academic connections, we used two datasets, Cora [10] and Citeseer [11], which are discussed in more detail in Section 3. In addition to network structure, these datasets contain class labels, relating to the publication venues, which can be used to test machine learning prediction algorithms. In datasets that exhibit homophily, utilizing the features of nodes within a topological arrangement (an adjacency matrix) produces improved classification performance over an instance only framework. Our investigation focuses on whether the explicit addition of weak ties will assist the inference process since they provide assistance in other network based processes. For instance, Roux et al. [12] investigate whether expertise within groups can cross the 'boundaries' from communities which are cohesive due to strong-tie connections. The work of [13] considers how differentiation of the edge types can improve the accuracy of social recommendations.

Determining the effect of interactions between nodes can be a time consuming process, which requires computational resources to analyze large networks. With the goal of understanding how node connections influence labels, various methodologies, such as [14], have been proposed where nodes iteratively propagate information throughout the network until convergence is achieved. A notable example of such is DeepWalk [15] which uses local information from truncated random walks to learn the latent variable representations. Relational neighbor classifiers such as Social Context Relational Neighbor (SCRN) have been shown to achieve good performance at inferring the labels of citation networks [14]. Graph Convolutional Networks (GCNs) [16] extended the methodology of Convolutional Neural Networks (CNNs) from images to graphs. GCNs, as CNNs, are constructed upon multiple layers of neural networks which makes them less amenable to interpretation [17–19].

This paper uses the approach of the Simplified Graph Convolutional Neural Networks (SGC) [20] to investigate the importance of strong vs. weak ties. The methodological framework, discussed in more detail in Section 4, provides a reduction in the complexity of the model and computation time required. It has an intuitive manner of producing feature projections and generating the non-linearity for different classes. Even though it is a deep learning approach that accounts for multiple-layers, the simplification allows a single parameter matrix to be produced that can more easily be interpreted if desired. An SGC implementation can be found in the DGL library [21,22].

The SGC uses the features of the nodes and the connectivity in the adjacency matrix to infer the class labels. In this paper we augment this adjacency matrix so that *weak-ties* are included as well. This produces a matrix in which the *strong-ties* and the *weak-ties* are treated equally at the start of the inference procedure. Results using this augmented adjacency matrix are compared to the label produced using the original adjacency matrix. Our experiments also consider the possibility of missing nodes. Obtaining complete datasets of networks is a challenge for a wide range of reasons; for instance, online platforms limit the API calls from developer accounts to reduce the website loads. It is then a crucial question as to whether the results of the investigation are sensitive to missing nodes [23]. Therefore, our experiments remove a range of pre-selected percentages of the network to compare the results. Nodes are ranked for removal based on three different centrality algorithms: betweenness, closeness, and VoteRank [24]. These algorithms sort the nodes in descending order and remove the top percentage chosen (e.g., 20%) so that the inference is performed without the influence of these nodes. We seek to answer the question: which gaps in the data are most likely to affect the SGC? The results help provide another piece of evidence towards the utility of *weak-ties* in sociological processes.

An added incentive for exploring the use of the SGC, is that it addresses an issue with the application of GCNs (graph convolutional neural networks) [16], where the increase in the number of layers beyond 2–3 can produce a degradation in the results. The number of layers employed by the GCN also corresponds to the K^{th} order neighborhood used in the SGC, and the results will be compared between both methodologies in Section 5. Although the application with GCNs [16] displays the degradation with an increase in the number of layers, L (corresponding to the K^{th} order neighborhood), the SGC does not display this degradation with an increase in K. The authors of [20] describe how the non-linearity for applications such as social networks may introduce unnecessary complexity.

The next section presents key work in the development of graph CNNs and other scientific explorations attesting to the power of weak-ties. Section 3 describes the datasets used in this study, and Section 4 outlines the methodology of the SGC. Then, in Section 5, we present results on the effects of augmenting the adjacency matrix with weak ties and removing nodes ranked on a selection of centrality measures. Within the results section is a subsection which compares the performance of the SGC with that of the GCN. Lastly, the conclusion is presented in Section 6, which summarizes the outcomes, outlines future work, and discusses the application to other datasets.

2. Related Work

This section presents an overview of related work on graph convolutional neural networks and weak ties.

2.1. Graph Convolutional Neural Networks

The work of [25] introduces how graph based methods can be used with convolutional neural networks (CNNs). These graph based methods are spectrally defined and their use within a spatial application utilizes recursive polynomials on the graph Laplacian. This enables spectrally motivated approaches to handle heterogeneous graphs. Convolutional neural networks [26] are employed as they allow for containing an efficient model architecture for extracting meaningful statistical patterns in high-dimensional datasets that can also be large (big data applications [27]). The application of CNNs to learn local stationary structures and apply them to hierarchical pattern searches has driven many advancements in ML tasks [28]. A key contribution of [25] is that the extension of the model to generalize to graphs is founded upon localized graph filters instead of the CNN localized convolution filter (or kernel). It presents a spectral graph formulation and how filters can be defined in respect to individual nodes in the graph with a certain number of 'hops' distance. An introduction to the field can be found in [29], where the reader can find a motivation for the fundamental analysis operations of signals from regular grids (lattice structures) to more general graphs. The authors of [30] build upon the work of signals on graphs and shows that a shift-invariant convolution filter can be formulated as a polynomial of adjacency matrices. These filters are defined as polynomials of functions of the graph

adjacency matrix, which describes an intuitive spatial formulation of the graph convolutional neural network. The use of the adjacency matrix is utilized in the approach described in Section 4. The filter uses the adjacency matrix and defines the exponent of the matrix as the degree polynomial. Effectively these exponents in the polynomial represent the number of edges ('hops') from any node.

The approach used here follows that of the work of [31], which relies on the adjacency matrix for filtering. It is noteworthy to emphasize that the graph based approach, which the authors provide code for, shows performance similar to the approach of CNNs the on CIFAR-10 and ImageNet datasets. This generalization could help understand how signals produced in the datasets can be collected whether they be image, document, sound based, brain region based, and more. Allowing for graph data that is heterogeneous is a flexibility which can produce more interesting applications. The model employed describes taking the single hop (defined edges) and the '2-hop' edges to be filtered upon that allows an extended radius of feature influence to be introduced for classification. This concept of using the adjacency matrix powers is explored in Section 5 where the exponent is taken over a range of values which represents the number of 'hops'. The authors of [16] also use this concept of the hop number that relies on the number of hidden-layers in the neural network and is the basis for the comparison approach employed in this work (described in Section 4).

Weak-Ties

After the publication of Granovetter's seminal work on the importance of weak ties [8] there have been multiple follow-up studies exploring how social links affect a member's ability to interact with others in the network and how different types of edges serve disparate roles in transmission and collaboration. Weak ties are important in many types of organizational structures; for instance, Patacchini et al. [32] studied the role of weak ties in criminal collusion. Since innovation requires teamwork and collaboration, organizations need to empower their workers to leverage their weak ties as described in [33]. The work of [34] looks at the effect of weak-ties on the job search process. Extracting information from the network of interactions is not a straightforward process and the work of [35] examines how this can be performed. Given recent evidence for increased social contention [36,37], the research of [38] considers the important question of how weak-ties can facilitate the increase of emotions such as anger on social media. Although weak-ties can be found to play a role in negative situations, there are other contexts where they play an important positive role, such as psychological well being where casual friendships add to happiness (strong ties plus weak ties) [39].

3. Data

Two datasets, Cora [10] and Citeseer [11], were used in this study. The Cora dataset is a citation network where the nodes refer to unique authors and the edges represent a weighted value for the mean citation relationship (from scientific publications). These scientific publications are classified and labeled into seven categories. The data were divided into a separate training and test set in order to provide consistent benchmarks between methodologies. The Citeseer dataset is another network dataset based upon citations where the nodes are also authors and edge values represent the mean citation relationship; it includes six different class labels. The SGC methodology described in Section 4 was used to infer the correct labels for a subset of the nodes in these datasets using both the original connectivity matrix and an augmented one. Table 1 provides an overview of some of the basic information of the datasets.

Figure 2 shows the degree distribution for the Cora dataset and how the distribution changes when different percentages of the network were removed. Those percentages of the network were removed according to different network centrality measures: betweenness in Figure 2a, closeness in Figure 2b, and VoteRank in Figure 2c. Figure 3 shows the same operation but using the Citeseer dataset. It is interesting to note how the Cora and Citeseer plots differ between their equivalent subfigures. The plots for the betweenness and the closeness change much more than VoteRank which provides

evidence that it is more robust against choosing nodes with many edges as a measure of centrality in different datasets.

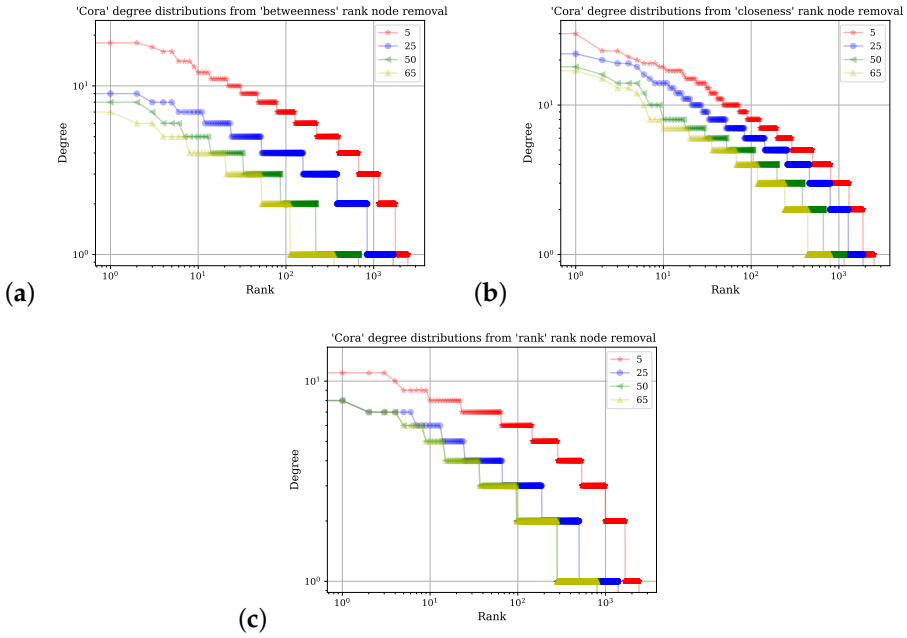

Figure 2. These plots show the degree distributions for the Cora network of publications [10] and how those distributions are altered when a certain percentage of the nodes are removed based upon a metric. Each subfigure shows the results of applying a different metric to sort and remove nodes: (**a**) node 'closeness'; (**b**) node 'betweenness'; (**c**) node 'VoteRank' [24].

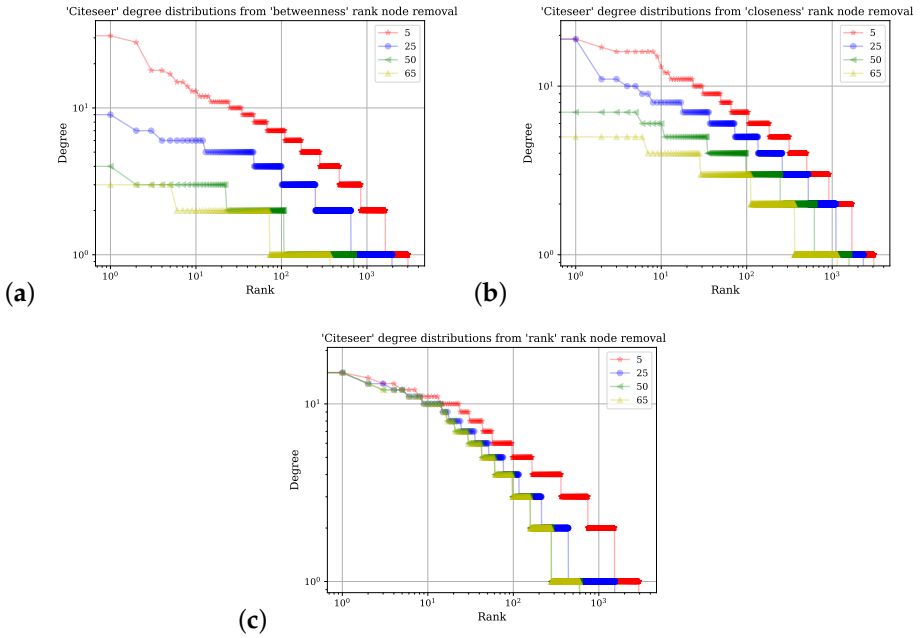

Figure 3. These plots show the degree distributions for the Citeseer network of publications [11] and how those distributions are altered when a certain percentage of the nodes are removed based upon a metric. Each subfigure shows the results of using a different metric to sort and remove nodes: (**a**) node 'closeness'; (**b**) node 'betweenness'; (**c**) node 'VoteRank' [24].

Table 1. Summary statistics of the networks from the datasets used in this study: Cora [10] and Citeseer [11]. Each of these datasets has a set of classes used to identify groups of publications in Cora as well as with Citeseer.

	Cora	**Citeseer**
# of Nodes	2708	3327
# of Edges	10,556	9228
# of Classes	7	6
Average degree	3.8981	2.8109
Density	0.00143	0.00084
Triadic closure	0.0934	0.13006

4. Methodology

For a graph $G = (V, \mathbf{A})$, $\mathcal{V}$ is the node within a set of N nodes $V = \{v_1, v_2, \ldots, v_N\}$, and the adjacency matrix is a symmetric matrix, $\mathbf{A} \in \mathbb{R}^{NxN}$. Each element of $\mathbf{A}$, $a_{i,j}$, holds the value of the weighted edge between two nodes v_i and v_j (an absence of an edge is represented by $a_{ij} = 0$). The degree matrix $\mathbf{D} = \text{diag}(d_1, d_2, \ldots, d_N)$ is a diagonal matrix of zero off diagonal entries and each diagonal entry is the row sum of the matrix $\mathbf{A}$; $d_i = \sum_j \mathbf{A}_{ij}$. There is a feature vector, $\mathbf{x}_i$, for each node i so that the set of features in the network of nodes is a $n \times d$ matrix, $\mathbf{X} \in \mathbb{R}^{n \times d}$ where d is the dimensionality of the feature vector. Each node is assigned a class label from the set of classes C; for each node we wish to utilize both $\mathbf{A}$ and $\mathbf{X}$ to infer $\mathbf{y}_i \in \{0,1\}^C$. $\mathbf{y}_i \in \{0,1\}^C$ is ideally a one-hot encoded vector which can be supplied data to assist the parameter estimations.

The normalized adjacency matrix with included self-loops is defined as,

$$\mathbf{S} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}, \tag{1}$$

and $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ and $\tilde{\mathbf{D}} = \text{diag}(\tilde{\mathbf{A}})$. The classifier employed by the SGC is:

$$\hat{Y} = \text{softmax}(\mathbf{S}^K \mathbf{X} \mathbf{\Theta}). \tag{2}$$

Here, the softmax can be replaced with σ as used in binary logistic regression when $C = 2$, and for the softmax on multiple categories we have $\text{softmax}(\mathbf{x}) = \exp(\mathbf{x}) / \sum_{c=1}^{C} \exp(\mathbf{x}_c)$. The component $\mathbf{\Theta}$ is the matrix of parameter values for the projections of the feature vectors so that it is of dimensionality $d \times C$, $\mathbf{\Theta} \in \mathbb{R}^{d \times C}$. Intuitively this can be understood as the parameter matrix holding a single vector of parameters of length equal to that of the feature vector and as many of these vectors as there are class labels. This linearization derives from the general concept in deep learning for sequential affine transforms in layers which are subsequent stages,

$$\hat{Y} = \text{softmax}(\mathbf{S} \ldots \mathbf{S}\mathbf{S}\mathbf{X}\mathbf{\Theta}^{(1)}\mathbf{\Theta}^{(2)} \ldots \mathbf{\Theta}^{(K)}). \tag{3}$$

It can then be seen how the value of K chosen represents the number of layers in the network employed. More details can be found in [20] where the methodological derivation is elaborated upon. A key requirement in this framework is the setting of the parameter value k. This can be considered as a tuning parameter for varying of the number of propagation steps taken. This relates to the matrix powers of an adjacency matrix which produce in each entry the number of 'walks' between nodes [40,41].

From the adjacency matrix the matrix including *weak-ties* produced through 'triangulation' ([9]) can be found via the walks of length two with $\mathbf{A}^2$. The original adjacency matrix is said to contain the *strong-ties* and there is considerable sociological research into the value of each type of connectivity [8]. In this work we explore the use of an adjacency matrix which contains both the *strong-ties* and the *weak-ties* via;

$$\mathbf{A}' = \mathbf{A}^2 + \mathbf{A}. \tag{4}$$

Figure 4 demonstrates this, and it can be seen visually in the subfigures. Figure 4a shows a hypothetical network with 4 nodes connected in a chain and Figure 4b shows how those nodes are connected when $\mathbf{A}'$ is produced from including both the *strong-ties* and the *weak-ties*.

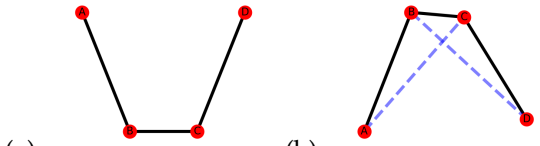

Figure 4. These show the concept of *weak-ties* produced through triangulation and how it can affect a small network: (**a**) shows a hypothetical network; (**b**) shows the result of introducing the *weak-ties* into the network as well as the original strong ties which were direct links.

Figure 5 shows a demonstration of the SGC methodology in its ability to accurately predict class labels on the Cora and Citeseer datasets. To explore how robust the methodology is, different percentages of the network were removed; nodes were selected for removal based on their rank calculated from different centrality measures: betweenness, closeness, and VoteRank. Each network measure expresses different aspects of a node's position in a network and therefore changes in the prediction accuracy, which assist in understanding empirically how node network placements contribute the most in correct label prediction. The VoteRank algorithm considers local node influences more than betweenness or closeness. Figure 5a shows results obtained from running the model on the Cora dataset, and the Citeseer results are shown in Figure 5b.

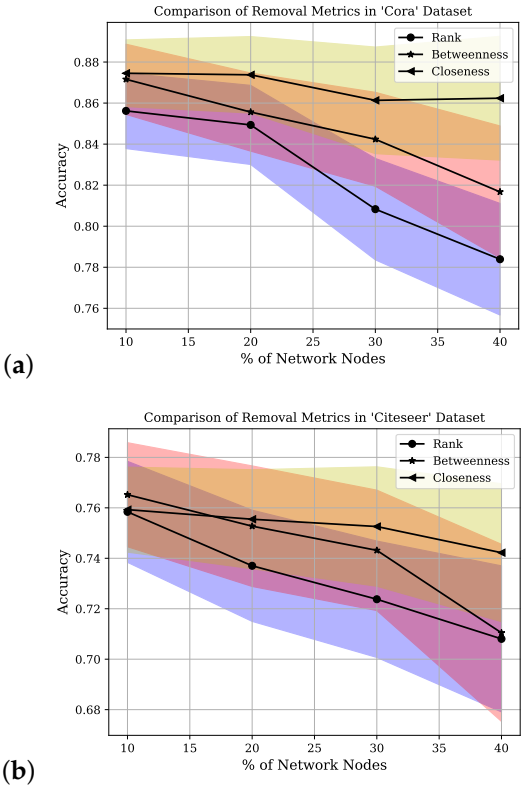

Figure 5. The Simplified Graph Convolutional Neural Networks (SGC) methodology was applied to predicting the test case labels when a certain percentage of the nodes were removed based upon the metrics closeness, betweenness, and VoteRank with the accuracy on the y-axis shown. These results shown in (**a**) and (**b**) are for the network datasets Citeseer and Cora.

5. Results

This section explores how the class label prediction accuracy is affected by different removal strategies when the connectivity matrix contains both the links for the strong ties and the weak ties. These results show how the parameter k can affect the accuracy of the prediction of class labels. Section 4 explores how the Simplified Graph Convolutional Neural Network (SGC) methodology performs on the datasets of Cora and Citeseer when different percentages of the nodes are removed. The nodes are removed according to their rank in terms of network centrality positions: betweenness, closeness, and VoteRank. For example, if 20% are removed using closeness as a measure, the nodes were ordered according to the value of closeness from largest to smallest, and the top 20% of the nodes in that percentile of closeness are removed. The purpose of this manipulation is to explore how robust the methodology is to central node removals whose influence on class labels can extend beyond their immediate vicinity.

As shown in Figure 5, we explore how the accuracy is affected by the different network measures used to rank nodes for removal but with the modified adjacency matrix that defines the connectivity for each node. This modification incorporates direct edges (links) called 'strong ties', as well as links between nodes that have a common friend. These newly introduced edges are the 'weak ties' that are a result of 'triangulation' as shown in Figure 4. The changes in the results due to the inclusion of the *weak-ties* can assist in establishing their importance in people's classification efforts in real life. A set of plots compare the accuracy of the SGC prediction of class labels with different network removal rankings given the addition of weak ties. The effect of the parameter value of k on accuracy is also explored to understand the sensitivity of the results to the only parameter that requires tuning in SGC.

Figure 6 shows the results of applying the SGC with different values of k for predicting class labels. On the horizontal axis is the value of k and on the vertical axis the accuracy as a percentage of the test class labels predicted correctly. The betweenness metric is used to rank the nodes and different percentages of the network's nodes are removed. The percentage values for each line are indicated in the legend. Figure 6a,b shows the results obtained from using the Cora and Citeseer datasets where the adjacency matrix used contains direct links between nodes and their *strong-ties* as well as their *weak-ties* as described in Section 4. Figure 6c,d shows the results when the original adjacency matrix containing only the *strong-ties* is used. For $k = 0$ similar results are obtained and for the final k value, $k = 7$, but the progression differs. The difference in progression is evident for the Cora dataset at $k = 1$ and up to $k = 4$ where the predictive accuracy for Figure 6a is reduced. This also applies to the Citeseer dataset, and especially to the scenario where 20% or 30% of the nodes have been removed. When $k = 0$ the SGC operates effectively in a manner similar to logistic regression where the network information is not used and inference is conducted using only the features of the node in question. These results support the conclusion that the augmented network topology of the strong-ties and the weak-ties does not facilitate improved accuracy of label prediction.

Figure 7 also shows the results of applying the SGC with different values of k for predicting class labels. The value of k is shown on the horizontal axis and on the vertical axis the accuracy as a percentage of test class labels being predicted correctly. Here the closeness metric is used to rank the nodes for removal. The different percentages for the removal of network nodes for each line is shown in the plot legends. Figure 7a,b shows the results obtained from using the Cora and Citeseer datasets where the adjacency matrix used contains direct links between nodes and their immediate neighbors (*strong-ties*) as well as their *weak-ties* (edges obtained via triangulation) as described in Section 4. Figure 7c,d shows the results when the original adjacency matrix containing only the *strong-ties* is used. For $k = 0$ similar results are obtained between the different pairs as the connectivity of the adjacency is not incorporated and node inference looks only at the features obtained from the node of concern. For $k = 7$ similar values are obtained through the extended radius of the adjacency power, but the progression of the trace differs between pairs of the plots. The difference between the pairs of traces can be easily seen by inspection of the application to the Cora dataset at values $k = 1$ and up to $k = 4$ where the predictive accuracy for Figure 7a is reduced. This also applies to the Citeseer dataset,

and is attenuated when 20% or 30% of the nodes have been removed. These results also support the conclusion that the augmented network topology of the strong-ties and the weak-ties does not facilitate improved accuracy of label prediction and that these conclusions are robust according to removal with a different network centrality ranking.

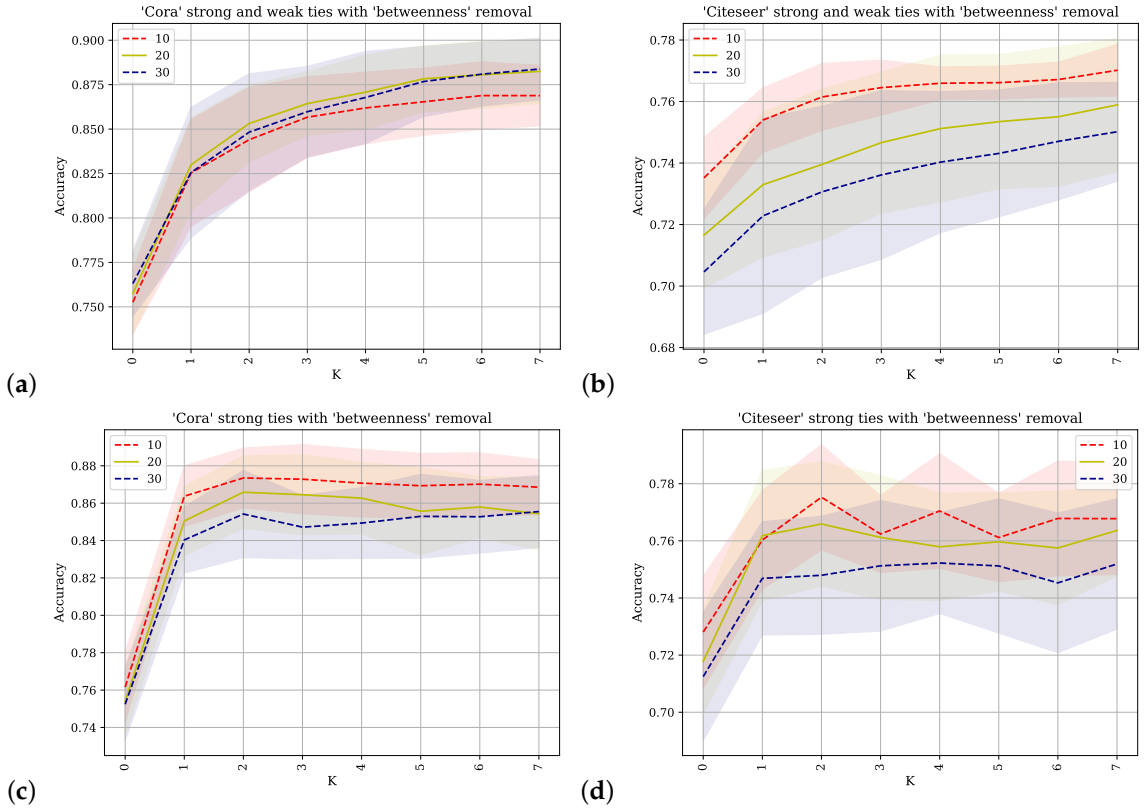

Figure 6. The SGC methodology was applied to predicting the class labels of the datasets Cora and Citeseer where the accuracy was plotted against the parameter k. The betweenness metric is used to rank and remove different percentages of the network: (**a**) and (**b**) show how the prediction changes when the network consists of *strong-ties* and *weak-ties*, and (**c**) and (**d**) show the results when the original adjacency matrix containing only *strong-ties* is used.

Figure 8 also shows the results of applying the SGC with different values of k for predicting class labels but uses the VoteRank centrality metric to rank the nodes for removal. The different percentages of node removal for each line are shown in the plot legends. Figure 8a,b shows the results obtained from the Cora and Citeseer datasets where the adjacency matrix used contains direct links between nodes and their immediate neighbors (*strong-ties*) as well as their *weak-ties* (edges obtained via triangulation) as described in Section 4. Figure 8c,d shows the results when the original adjacency matrix containing only the *strong-ties* is used. When $k = 0$ similar results are obtained between the different pairs as the connectivity of the adjacency is not incorporated and node inference looks only at the features from the node of concern. The application of VoteRank changes the interpretation of the previous results where both applications to Cora and Citeseer have improved results for the augmented adjacency matrix (*strong-ties* and *weak-ties*) from $k = 3$ and upwards.

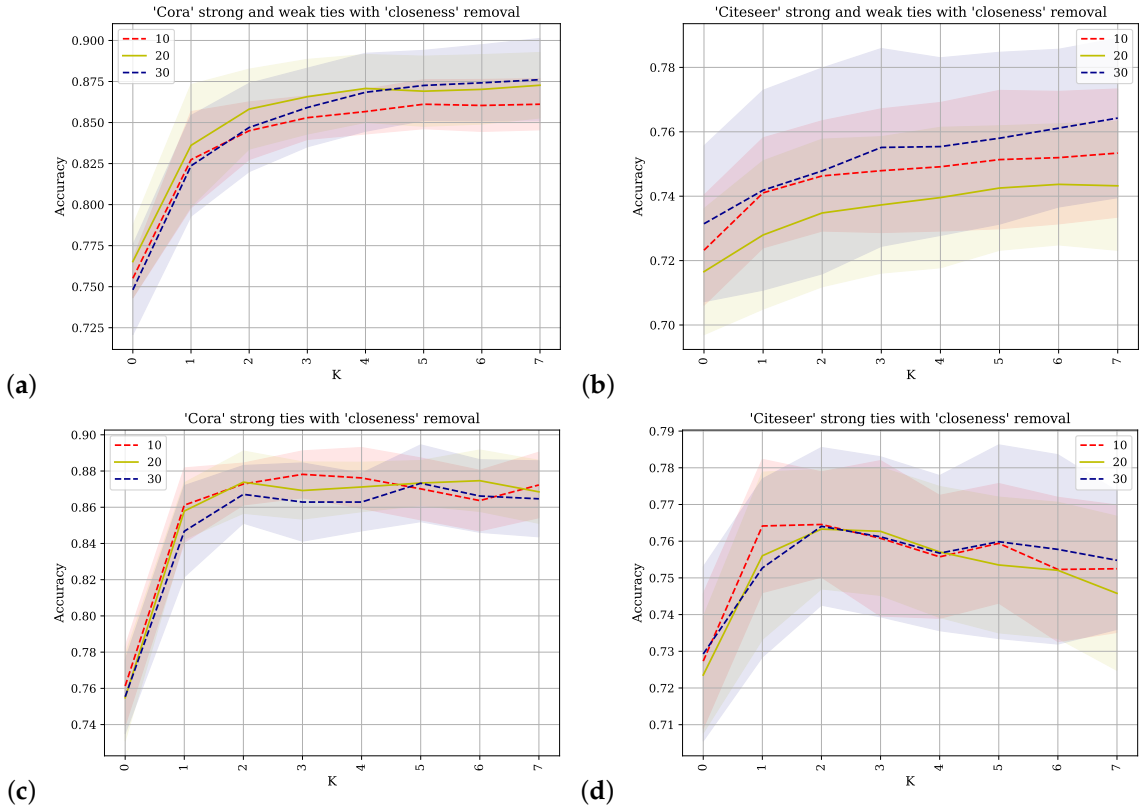

Figure 7. The SGC methodology was applied to predicting the class labels of the Cora and Citeseer datasets where the accuracy was plotted against the parameter k. The closeness metric is used to rank and remove different percentages of the network: (**a**) and (**b**) show the prediction changes when the network consists of *strong-ties* and *weak-ties*, and (**c**) and (**d**) show the results when the original adjacency matrix containing only *strong-ties* is used.

The set of results show that for $k < 3$ the adjacency matrix containing the set of original *strong-ties* edges suffices to produce the best results. For larger values of k the augmented adjacency, which contains both the *strong-ties* and the *weak-ties*, can show improved performance when nodes are removed according to the VoteRank algorithm and not according to betweenness or closeness. This emphasizes that there is a complex interplay between how node centrality is measured and the manner in which the inference methodology operates. It cannot be considered an *a priori* principle that *weak-ties* can provide an increased predictive power due to its support from the social science domain and its adherence to it. On the contrary, they induce a requirement for larger values of k to reach the maximum accuracy implying that the SGC requires more 'layers' which effectively aggregates information from more distant nodes in order to counter balance the introduction of *weak-ties* as *strong-ties*. This can provide anecdotal evidence that those two types of edges may require separate treatment. Further experiments conducted, working with a starting network of only the *weak-ties*, produced networks with an increased number of disconnected components.

These results also support the claims of the authors of 'VoteRank' when they state that the methodology identifies a set of decentralized spreaders as opposed to focusing on a group of spreaders which overlap in their sphere of influence. This is why the VoteRank targeted node removal was more effective in reducing the accurate label inference since more locally influential nodes for classification were identified; the *weak-ties* provided extra information about local labels in the absence of these essential *strong-tie* connected nodes.

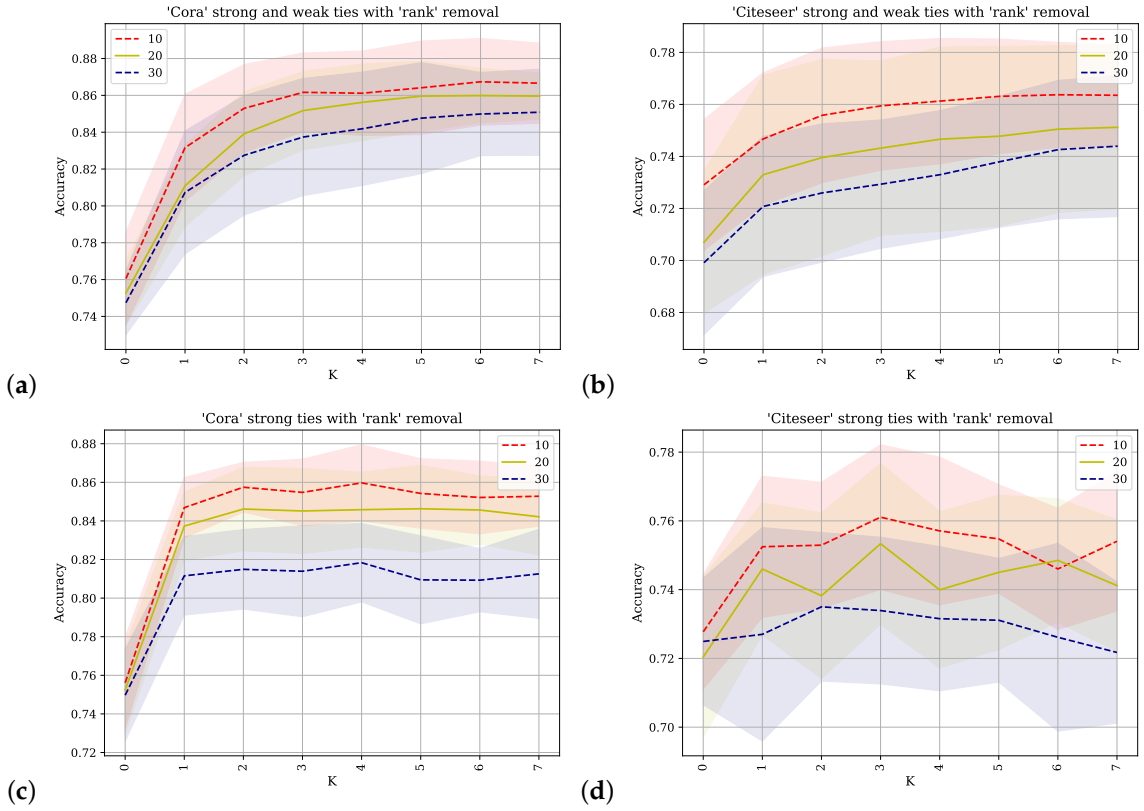

Figure 8. The SGC methodology was applied to predicting the class labels of the Cora and Citeseer datasets where the accuracy is plotted against the parameter k. The VoteRank metric is used to rank and remove different percentages of the network: (**a**) and (**b**) show the prediction changes when the network consists of *strong-ties* and *weak-ties*, and (**c**) and (**d**) show the results when the original adjacency matrix containing only *strong-ties* is used.

Comparison to GCN

This section compares results from applying SGC [20] vs. using the original GCN framework [16]. Appendix B of [16] discusses the effect of adding more network layers on accuracy. It states that the best choice is 2–3 layers and that after 7 layers there is steep degradation of accuracy. The number of layers corresponds to the number of 'K' hops as explored with the SGC previously. The SGC methodology encapsulates the K hop neighborhood without the non-linearity and therefore avoids the degradation of accuracy with increased K or L. Figures 9–11 present the results of applying the GCN in the same set of situations that we evaluated with SGC. The number of layers L is on the x-axis (corresponding analogously to the K in SGC) and the y-axis is the accuracy. In each of these figures, the Cora and Citeseer datasets are used when examining the strong with weak ties in an augmented network as well as using the original network containing only the strong ties. Each figure removes a percentage of the nodes based upon the rank of the nodes with the centrality measures of betweenness, closeness, and VoteRank respectively. The plots have three lines per plot where there are different percentages (10%, 20%, and 30%) of the nodes removed based upon the centrality measure. In each of the Figures 9–11 it can be seen how the accuracy degrades after $L = 1$ showing how the SGC is able to include more network information about each node without introducing unnecessary complexity which degrades accuracy. The degradation of the accuracy in relation to the choice of centrality measure is comparable between the results, showing that the GCN is less specific to the node network positions than the SGC is, which can be attributed to the non-linearity the GCN introduces via the layers.

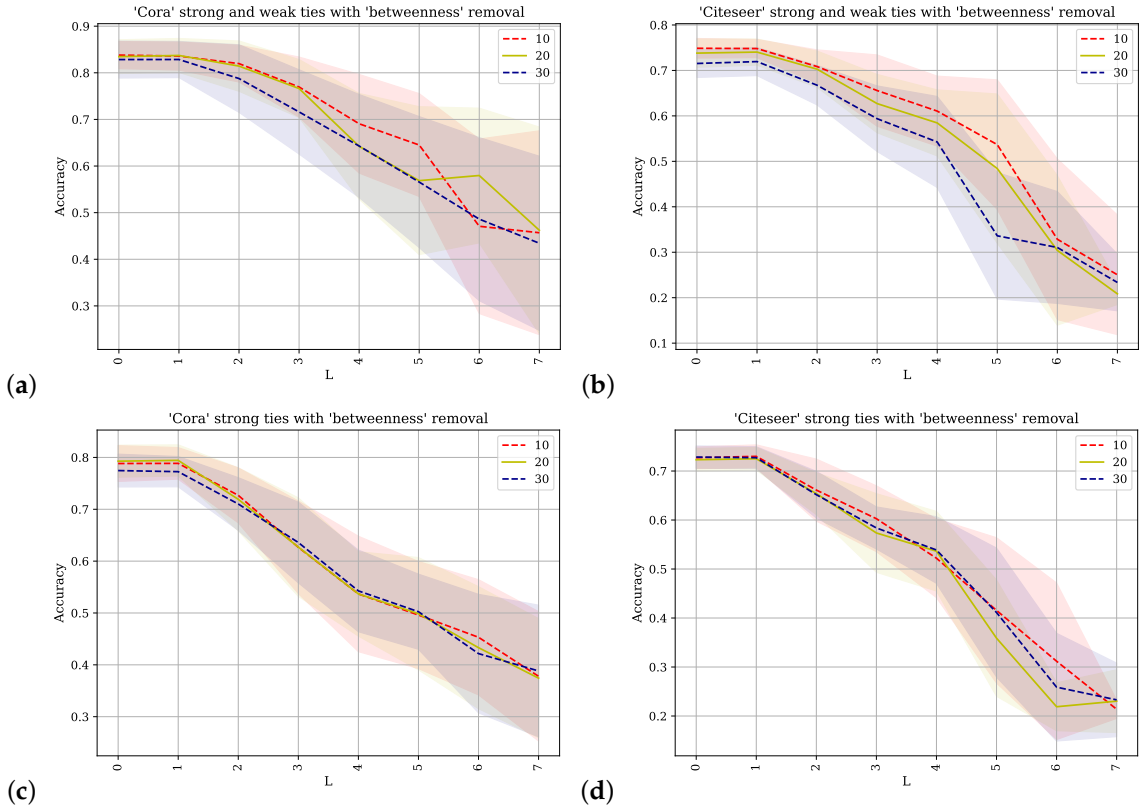

Figure 9. The GCN methodology was applied to predicting the class labels of the datasets Cora and Citeseer where the accuracy is plotted against the parameter L. The betweenness metric was used to rank and remove different percentages of the network: (**a**) and (**b**) show the prediction changes when the network consists of *strong-ties* and *weak-ties*, and (**c**) and (**d**) show the results when the original adjacency matrix containing only *strong-ties* is used.

In Appendix A, an additional set of tables are provided in order to see the comparison between the effectiveness of the SGC and the GCN over the values of K and L respectively. The two datasets and the centrality measures with the different edge sets are examined.

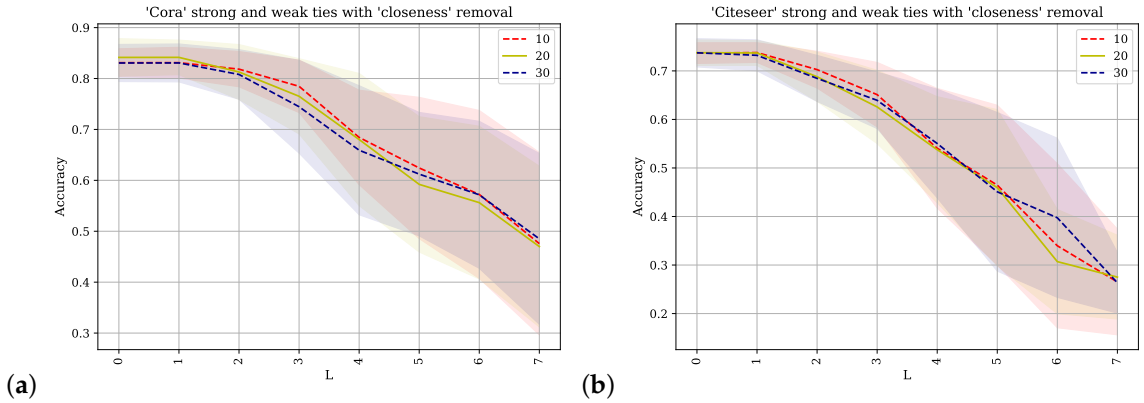

Figure 10. *Cont.*

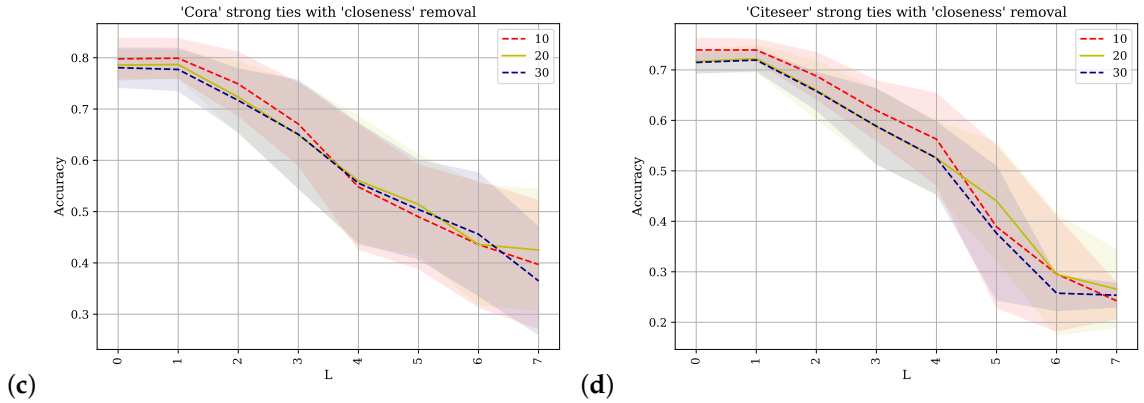

Figure 10. The Graph Convolutional Networks (GCNs) methodology was applied to predicting the class labels of the Cora and Citeseer datasets where the accuracy is plotted against the parameter l. The closeness metric is used to rank and remove different percentages of the network: (**a**) and (**b**) show the prediction changes when the network consists of *strong-ties* and *weak-ties*, and (**c**) and (**d**) show the results when the original adjacency matrix containing only *strong-ties* is used.

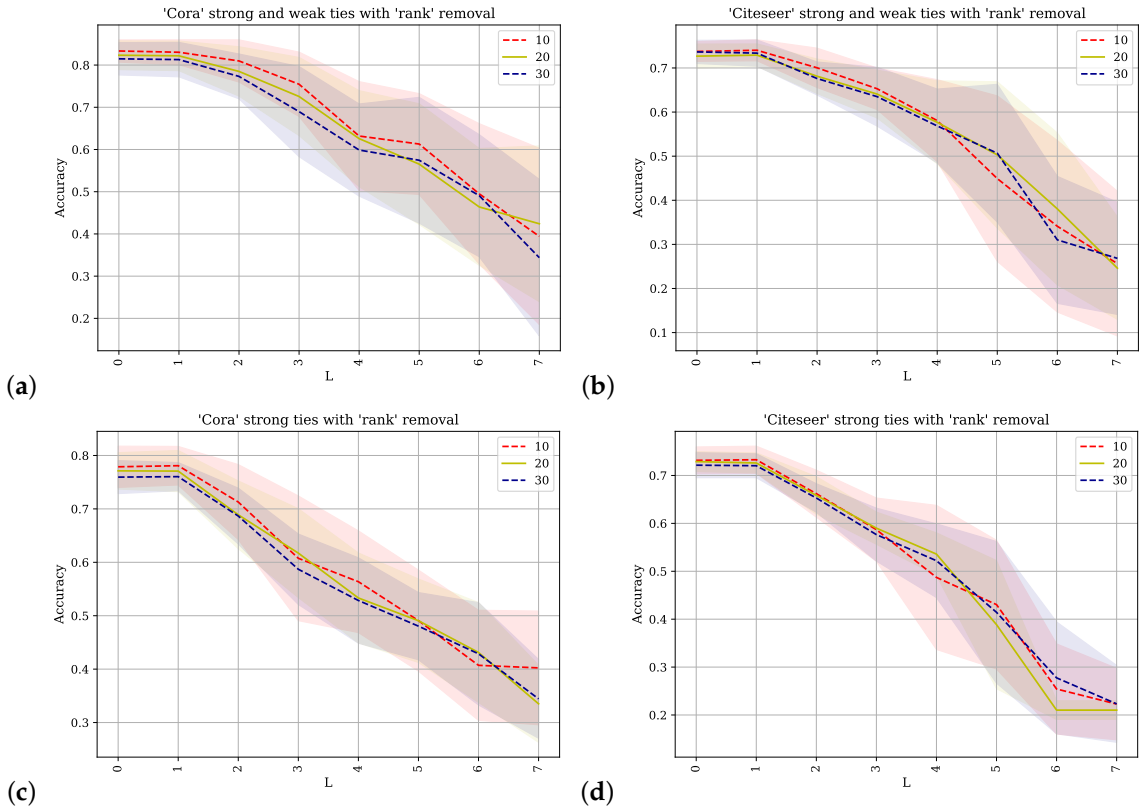

Figure 11. The GCN methodology was applied to predicting the class labels of the Cora and Citeseer datasets where the accuracy is plotted against the parameter l. The VoteRank metric is used to rank and remove different percentages of the network: (**a**) and (**b**) show the prediction changes when the network consists of *strong-ties* and *weak-ties*, and (**c**) and (**d**) show the results when the original adjacency matrix containing only *strong-ties* is used.

6. Conclusions

This paper explores the uses of the recently introduced methodology, the Simplified Graph Convolutional Neural Network (SGC); class label inferences are produced based on the network structure, represented by an adjacency matrix, in combination with node feature vectors. There is interest in exploring this model in more depth since it provides a succinct yet expressive formulation

for describing how nodes can influence class label prediction within a network. Besides the parameters fitted in order to optimize the target label prediction, there is only a single parameter value k, which requires manual tuning. This parameter is related to the number of layers S^k (described in Section 4).

The exploration conducted here investigates the degree to which the accurate prediction of class labels is reduced by removing percentages of the network ranked by centrality metrics. This provides evidence to the practitioner who collects data, that may contain gaps in the network, and needs to know if the conclusions can be drastically affected by missing data on key nodes as to whether the the SGC is sensitive to such issues. Three different network centrality measures are used to select removal nodes: betweenness, closeness, and VoteRank. We find that the methodology does manage to produce analogous predictions based upon different percentages of removal (10/20/30). The largest observed changes were when the nodes were selected for removal with the VoteRank algorithm and not with betweenness or closeness. This shows that the SGC label assignments are more sensitive to the local label information derived from the features of the local nodes than well connected groups of nodes in the center of the network. This also explains why it has displayed the ability to be robust in its predictions.

The other question explored is whether the results would change if the SGC was supplied an adjacency matrix that contained the 'triangulated edges' to begin with. The existing edges in the adjacency matrix can be referred to as *strong-ties* as they are direct links; the edges that connect friends-of-friends (produced from triangulation A^2), can be referred to as *weak-ties*. A matrix with both of these edge sets was supplied to the SGC to compare the accuracy predictions. There is considerable sociological literature discussing the importance of these edges in helping to discover important connections. Our results show a degraded outcome with the exception of when nodes were removed with the VoteRank algorithm. This indicates that the inclusion of the *weak-ties* provides a more robust edge set when important local nodes are removed. The results do not show an ability to improve the prediction of class labels for low removal percentages when *weak-ties* are included.

The datasets used in this study contained monolithic graphs, where every node is reachable from any other node. There are many datasets where the data contains disjoint graphs, and this can be particularly common when the observational capabilities are limited in comparison to the process. A notable example is with protein interaction graphs. Applying the investigation taken here with such data would alter the adjacency matrix but not in a way that would cause a failure in its ability to follow the procedures described. Since the exploration did not depend upon a small fraction of the number of nodes, the study could continue with such data as long as the distribution of the relative betweenness and closeness is not excessively skewed for the subgraphs. The investigation therefore can be conducted on a wide range of datasets to explore the role of weak ties in the networks. Corporate networks are an interesting avenue for extensions as the nodes would be more 'complex' entities which may rely on their network connections in different ways. A key aspect of the extendibility is the overhead of the approach. Since the parameter, feature and adjacency matrix are combined with linear operators with a non-nested set of intermediate features, inferences are relatively cheaper than other approaches that build deeper trees and introduce further non-linearities.

Author Contributions: Conceptualization, N.H.B. and A.V.M.; formal analysis, N.H.B. and A.V.M.; investigation, N.H.B. and A.V.M.; methodology, N.H.B. and A.V.M.; resources, G.S.; software, N.H.B.; supervision, A.V.M. and G.S.; validation, N.H.B. and A.V.M.; visualization, N.H.B.; writing—original draft, N.H.B.; writing—review and editing, A.V.M. and G.S. All authors have read and agreed to the published version of the manuscript.

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.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Comparison of Results between the SGC and GCN

The following tables present the comparison of the results between the use of the SGC and the GCN (shown separately in the Results section). Each table lists the centrality metric used to remove nodes based upon its ranking: betweenness, closeness and VoteRank. The column 'P' identifies the percentage of the network nodes removed based upon that centrality metric. The column 'L' represents the number of layers used by the GCN and the column 'K' is for the exponent of the normalized adjacency matrix in the SGC. Under the columns 'GCN' and 'SGC' which refer to the graph convolutional neural network and simplified graph convolutional neural network respectively are the columns 'S' for the strong-ties used and 'SW' for the strong-ties and weak-ties aggregated. In Tables A1–A6, the cell entries show the accuracy; each centrality measure is reported for the two datasets Cora and Citeseer.

Table A1. The GCN and SGC methodologies were applied to predicting the class labels of the Cora dataset. The betweenness metric is used to rank and remove different percentages of the network. *S* stands for when the network has the initial strong connections only and SW represents when the network is augmented with weak ties alongside the strong ties. *L* and *K* denotes the number of layers and the power in GCN and SGC framework respectively.

Metric	P	GCN					SGC			
		L	S		SW		K	S		SW	
			Mean	std	Mean	std		Mean	std	Mean	std
Betweenness	10	0	0.79	0.03	0.84	0.03	0	0.76	0.02	0.74	0.01
		1	0.79	0.03	0.84	0.03	1	0.86	0.02	0.75	0.01
		2	0.73	0.05	0.82	0.04	2	0.88	0.01	0.76	0.01
		3	0.63	0.08	0.77	0.06	3	0.87	0.02	0.76	0.01
		4	0.54	0.08	0.69	0.11	4	0.87	0.02	0.77	0.01
		5	0.50	0.07	0.65	0.11	5	0.87	0.02	0.77	0.01
		6	0.45	0.11	0.47	0.19	6	0.87	0.02	0.77	0.01
		7	0.38	0.13	0.46	0.22	7	0.87	0.01	0.77	0.01
	20	0	0.79	0.03	0.83	0.04	0	0.76	0.02	0.72	0.02
		1	0.79	0.03	0.84	0.04	1	0.85	0.02	0.73	0.02
		2	0.72	0.06	0.81	0.05	2	0.86	0.02	0.74	0.02
		3	0.63	0.09	0.77	0.06	3	0.86	0.02	0.75	0.02
		4	0.54	0.08	0.64	0.11	4	0.86	0.02	0.75	0.02
		5	0.50	0.11	0.57	0.16	5	0.86	0.02	0.75	0.02
		6	0.43	0.12	0.58	0.14	6	0.86	0.02	0.76	0.02
		7	0.37	0.11	0.46	0.22	7	0.85	0.02	0.76	0.02
	30	0	0.77	0.03	0.83	0.04	0	0.76	0.02	0.70	0.02
		1	0.77	0.03	0.83	0.04	1	0.85	0.02	0.72	0.03
		2	0.71	0.05	0.79	0.07	2	0.85	0.02	0.73	0.03
		3	0.64	0.09	0.72	0.09	3	0.85	0.02	0.74	0.03
		4	0.54	0.11	0.64	0.11	4	0.85	0.02	0.74	0.02
		5	0.50	0.10	0.57	0.14	5	0.85	0.02	0.74	0.02
		6	0.42	0.11	0.49	0.17	6	0.85	0.02	0.75	0.02
		7	0.39	0.12	0.43	0.19	7	0.86	0.02	0.75	0.02

Table A2. The GCN and SGC methodologies were applied to predicting the class labels of the Cora dataset. The closeness metric is used to rank and remove different percentages of the network. *S* stands for when the network has the initial strong connections only and SW represents when network is augmented with weak ties alongside the strong ties. *L* and *K* denotes the number of layers and the power in GCN and SGC framework respectively.

Metric	**P**	**GCN**					**SGC**			
		L	**S**		**SW**		**K**	**S**		**SW**	
			Mean	**std**	**Mean**	**std**		**Mean**	**std**	**Mean**	**std**
Closeness	10	0	0.80	0.04	0.83	0.03	0	0.77	0.02	0.76	0.01
		1	0.80	0.04	0.83	0.03	1	0.85	0.02	0.83	0.03
		2	0.75	0.06	0.82	0.03	2	0.87	0.01	0.85	0.02
		3	0.67	0.08	0.78	0.05	3	0.87	0.01	0.85	0.01
		4	0.55	0.12	0.68	0.09	4	0.87	0.02	0.86	0.01
		5	0.49	0.10	0.62	0.14	5	0.88	0.01	0.86	0.01
		6	0.44	0.12	0.57	0.16	6	0.86	0.02	0.86	0.02
		7	0.40	0.12	0.48	0.18	7	0.87	0.02	0.86	0.02
	20	0	0.79	0.03	0.84	0.04	0	0.76	0.02	0.77	0.02
		1	0.79	0.03	0.84	0.03	1	0.86	0.01	0.84	0.04
		2	0.72	0.07	0.81	0.05	2	0.88	0.02	0.86	0.02
		3	0.65	0.10	0.77	0.07	3	0.87	0.02	0.87	0.02
		4	0.56	0.13	0.68	0.13	4	0.88	0.02	0.87	0.02
		5	0.51	0.10	0.59	0.13	5	0.87	0.01	0.87	0.02
		6	0.44	0.12	0.56	0.15	6	0.87	0.02	0.87	0.02
		7	0.43	0.12	0.47	0.16	7	0.87	0.02	0.87	0.02
	30	0	0.78	0.04	0.83	0.04	0	0.75	0.02	0.75	0.03
		1	0.78	0.04	0.83	0.04	1	0.86	0.02	0.82	0.03
		2	0.72	0.06	0.81	0.05	2	0.87	0.02	0.85	0.03
		3	0.65	0.11	0.74	0.09	3	0.87	0.02	0.86	0.02
		4	0.56	0.12	0.66	0.13	4	0.87	0.03	0.87	0.02
		5	0.50	0.10	0.61	0.12	5	0.87	0.03	0.87	0.02
		6	0.46	0.12	0.57	0.14	6	0.87	0.02	0.87	0.02
		7	0.36	0.10	0.49	0.17	7	0.87	0.02	0.88	0.03

Table A3. The GCN and SGC methodologies were applied to predicting the class labels of the Cora dataset. The VoteRank metric is used to rank and remove different percentages of the network. *S* stands for when the network has the initial strong connections only and SW represents when network is augmented with weak ties alongside the strong ties. *L* and *K* denotes the number of layers and the power in GCN and SGC framework respectively.

Metric	**P**	**GCN**					**SGC**			
		L	**S**		**SW**		**K**	**S**		**SW**	
			Mean	**std**	**Mean**	**std**		**Mean**	**std**	**Mean**	**std**
Vote Rank	10	0	0.78	0.04	0.83	0.03	0	0.75	0.02	0.76	0.02
		1	0.78	0.04	0.83	0.03	1	0.86	0.01	0.83	0.03
		2	0.71	0.07	0.81	0.05	2	0.87	0.02	0.85	0.02
		3	0.61	0.12	0.75	0.08	3	0.87	0.02	0.86	0.02
		4	0.56	0.10	0.63	0.13	4	0.87	0.02	0.86	0.02
		5	0.49	0.09	0.61	0.12	5	0.87	0.01	0.86	0.03
		6	0.41	0.10	0.49	0.17	6	0.87	0.02	0.87	0.02
		7	0.40	0.11	0.39	0.21	7	0.87	0.02	0.87	0.02
	20	0	0.77	0.03	0.82	0.03	0	0.75	0.02	0.75	0.02
		1	0.77	0.04	0.82	0.04	1	0.85	0.02	0.81	0.02
		2	0.69	0.06	0.78	0.06	2	0.86	0.02	0.84	0.02
		3	0.62	0.08	0.73	0.09	3	0.86	0.02	0.85	0.02
		4	0.53	0.08	0.63	0.11	4	0.86	0.03	0.86	0.02
		5	0.49	0.08	0.57	0.14	5	0.85	0.03	0.86	0.02
		6	0.43	0.09	0.46	0.14	6	0.85	0.02	0.86	0.01
		7	0.34	0.07	0.42	0.18	7	0.85	0.02	0.86	0.01
	30	0	0.76	0.03	0.81	0.04	0	0.76	0.02	0.75	0.02
		1	0.76	0.03	0.81	0.04	1	0.84	0.02	0.81	0.03
		2	0.69	0.05	0.77	0.05	2	0.85	0.01	0.83	0.03
		3	0.59	0.07	0.69	0.11	3	0.84	0.02	0.84	0.03
		4	0.53	0.08	0.60	0.11	4	0.85	0.02	0.84	0.03
		5	0.48	0.06	0.57	0.15	5	0.85	0.02	0.85	0.03
		6	0.43	0.10	0.49	0.14	6	0.85	0.02	0.85	0.02
		7	0.34	0.07	0.34	0.19	7	0.84	0.01	0.85	0.02

Table A4. The GCN and SGC methodologies were applied to predicting the class labels of the Citeseer dataset. The betweenness metric is used to rank and remove different percentages of the network. *S* stands for when the network has the initial strong connections only and SW represents when network is augmented with weak ties alongside the strong ties. *L* and *K* denotes the number of layers and the power in GCN and SGC framework respectively.

Metric	P	GCN					SGC			
		L	S		SW		K	S		SW	
			Mean	std	Mean	std		Mean	std	Mean	std
Betweenness	10	0	0.73	0.02	0.75	0.02	0	0.72	0.02	0.74	0.01
		1	0.73	0.02	0.75	0.02	1	0.76	0.02	0.75	0.01
		2	0.66	0.06	0.71	0.04	2	0.78	0.02	0.76	0.01
		3	0.60	0.07	0.66	0.08	3	0.76	0.01	0.76	0.01
		4	0.52	0.08	0.61	0.08	4	0.77	0.01	0.77	0.01
		5	0.42	0.15	0.54	0.14	5	0.77	0.02	0.77	0.01
		6	0.31	0.16	0.33	0.18	6	0.76	0.02	0.77	0.01
		7	0.21	0.02	0.25	0.13	7	0.77	0.02	0.77	0.01
	20	0	0.72	0.02	0.74	0.03	0	0.71	0.01	0.72	0.02
		1	0.73	0.02	0.74	0.03	1	0.76	0.02	0.73	0.02
		2	0.65	0.04	0.70	0.04	2	0.76	0.02	0.74	0.02
		3	0.57	0.08	0.63	0.06	3	0.77	0.02	0.75	0.02
		4	0.54	0.08	0.58	0.07	4	0.75	0.02	0.75	0.02
		5	0.36	0.12	0.48	0.16	5	0.76	0.02	0.75	0.02
		6	0.22	0.05	0.30	0.16	6	0.76	0.02	0.76	0.02
		7	0.23	0.06	0.21	0.02	7	0.77	0.02	0.76	0.02
	30	0	0.73	0.02	0.72	0.03	0	0.71	0.02	0.70	0.02
		1	0.73	0.02	0.72	0.03	1	0.74	0.02	0.72	0.03
		2	0.65	0.05	0.67	0.04	2	0.75	0.02	0.73	0.03
		3	0.58	0.04	0.59	0.07	3	0.75	0.02	0.74	0.03
		4	0.54	0.07	0.54	0.10	4	0.75	0.01	0.74	0.02
		5	0.41	0.13	0.34	0.14	5	0.76	0.02	0.74	0.02
		6	0.26	0.11	0.31	0.12	6	0.74	0.03	0.75	0.02
		7	0.23	0.08	0.23	0.06	7	0.75	0.02	0.75	0.02

Table A5. The GCN and SGC methodologies were applied to predicting the class labels of the Citeseer dataset. The closeness metric is used to rank and remove different percentages of the network. *S* stands for when the network has the initial strong connections only and SW represents when network is augmented with weak ties alongside the strong ties. *L* and *K* denotes the number of layers and the power in GCN and SGC framework respectively.

Metric	P	GCN					SGC			
		L	S		SW		K	S		SW	
			Mean	std	Mean	std		Mean	std	Mean	std
Closeness	10	0	0.74	0.02	0.74	0.02	0	0.72	0.02	0.72	0.02
		1	0.74	0.02	0.74	0.02	1	0.76	0.02	0.74	0.02
		2	0.69	0.05	0.70	0.04	2	0.76	0.02	0.75	0.02
		3	0.62	0.06	0.65	0.07	3	0.76	0.01	0.75	0.02
		4	0.56	0.09	0.54	0.12	4	0.76	0.01	0.75	0.02
		5	0.39	0.16	0.46	0.16	5	0.76	0.02	0.75	0.02
		6	0.30	0.11	0.34	0.17	6	0.74	0.02	0.75	0.02
		7	0.24	0.04	0.27	0.11	7	0.74	0.02	0.75	0.02
	20	0	0.72	0.02	0.74	0.03	0	0.72	0.02	0.72	0.02
		1	0.72	0.03	0.74	0.03	1	0.76	0.03	0.73	0.02
		2	0.66	0.06	0.69	0.05	2	0.76	0.01	0.73	0.02
		3	0.59	0.07	0.63	0.08	3	0.77	0.02	0.74	0.02
		4	0.53	0.07	0.54	0.11	4	0.75	0.02	0.74	0.02
		5	0.44	0.12	0.46	0.16	5	0.75	0.01	0.74	0.02
		6	0.30	0.12	0.31	0.11	6	0.74	0.02	0.74	0.02
		7	0.27	0.08	0.28	0.09	7	0.74	0.03	0.74	0.02
	30	0	0.71	0.02	0.74	0.03	0	0.71	0.03	0.73	0.02
		1	0.72	0.02	0.73	0.03	1	0.75	0.02	0.74	0.03
		2	0.66	0.04	0.68	0.05	2	0.76	0.03	0.75	0.03
		3	0.59	0.07	0.64	0.06	3	0.77	0.01	0.76	0.03
		4	0.53	0.07	0.55	0.11	4	0.76	0.03	0.76	0.03
		5	0.38	0.13	0.45	0.16	5	0.75	0.02	0.76	0.03
		6	0.26	0.03	0.40	0.16	6	0.76	0.02	0.76	0.02
		7	0.25	0.02	0.26	0.06	7	0.76	0.02	0.76	0.02

Table A6. The GCN and SGC methodologies were applied to predicting the class labels of the Citeseer data set. The VoteRank metric is used to rank and remove different percentages of the network. *S* stands for when the network has the initial strong connections only and SW represents when network is augmented with weak ties alongside the strong ties. *L* and *K* denotes the number of layers and the power in GCN and SGC framework respectively.

Metric	P	GCN					SGC			
		L	S		SW		K	S		SW	
			Mean	std	Mean	std		Mean	std	Mean	std
Vote Rank	10	0	0.73	0.03	0.74	0.02	0	0.72	0.02	0.73	0.03
		1	0.73	0.03	0.74	0.02	1	0.76	0.01	0.75	0.03
		2	0.66	0.05	0.70	0.04	2	0.76	0.01	0.76	0.03
		3	0.59	0.07	0.65	0.05	3	0.74	0.01	0.76	0.02
		4	0.49	0.15	0.58	0.09	4	0.76	0.01	0.76	0.02
		5	0.43	0.13	0.45	0.19	5	0.75	0.02	0.76	0.02
		6	0.25	0.09	0.34	0.19	6	0.76	0.02	0.76	0.02
		7	0.22	0.07	0.26	0.16	7	0.75	0.02	0.76	0.02
	20	0	0.73	0.02	0.73	0.03	0	0.73	0.01	0.71	0.03
		1	0.73	0.02	0.73	0.03	1	0.71	0.01	0.73	0.04
		2	0.66	0.04	0.68	0.04	2	0.75	0.02	0.74	0.04
		3	0.59	0.03	0.64	0.05	3	0.75	0.01	0.74	0.03
		4	0.54	0.04	0.58	0.09	4	0.77	0.02	0.75	0.04
		5	0.39	0.13	0.50	0.17	5	0.75	0.02	0.75	0.03
		6	0.21	0.02	0.38	0.17	6	0.76	0.02	0.75	0.03
		7	0.21	0.02	0.25	0.12	7	0.75	0.04	0.75	0.03
	30	0	0.72	0.03	0.74	0.03	0	0.73	0.01	0.70	0.03
		1	0.72	0.02	0.73	0.03	1	0.71	0.03	0.72	0.03
		2	0.65	0.03	0.68	0.04	2	0.75	0.01	0.73	0.03
		3	0.58	0.06	0.63	0.07	3	0.72	0.03	0.73	0.02
		4	0.52	0.08	0.57	0.08	4	0.73	0.01	0.73	0.02
		5	0.41	0.15	0.51	0.16	5	0.74	0.03	0.74	0.03
		6	0.28	0.12	0.31	0.14	6	0.74	0.04	0.74	0.03
		7	0.22	0.08	0.27	0.13	7	0.74	0.02	0.74	0.03

References

1. Kwak, H.; Lee, C.; Park, H.; Moon, S. What is Twitter, a social network or a news media? In Proceedings of the International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 591–600.
2. Wadhwa, V.; Latimer, E.; Chatterjee, K.; McCarty, J.; Fitzgerald, R. Maximizing the tweet engagement rate in academia: Analysis of the AJNR Twitter feed. *Am. J. Neuroradiol.* **2017**, *38*, 1866–1868.
3. Shuai, X.; Pepe, A.; Bollen, J. How the scientific community reacts to newly submitted preprints: Article downloads, Twitter mentions, and citations. *PLoS ONE* **2012**, *7*, e47523.
4. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The Pagerank Citation Ranking: Bringing Order to the Web*; Technical Report; Stanford InfoLab: Stanford, CA, USA, 1999.
5. Ding, Y.; Yan, E.; Frazho, A.; Caverlee, J. PageRank for ranking authors in co-citation networks. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 2229–2243.

6. Ding, Y. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *J. Inf.* **2011**, *5*, 187–203.
7. Bian, Y. Bringing strong ties back in: Indirect ties, network bridges, and job searches in China. *Am. Sociol. Rev.* **1997**, *62*, 366–385.
8. Granovetter, M.S. The strength of weak ties. In *Social Networks*; Elsevier: Amsterdam, The Netherlands, 1977; pp. 347–367.
9. Mantzaris, A.V.; Higham, D.J. Infering and calibrating triadic closure in a dynamic network. In *Temporal Networks*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 265–282.
10. Mccallum, A. CORA Research Paper Classification Dataset. 2001. Available online: people.cs.umass.edu/mccallum/data.html.KDD (accessed on 20 May 2020).
11. Caragea, C.; Wu, J.; Ciobanu, A.; Williams, K.; Fernández-Ramírez, J.; Chen, H.H.; Wu, Z.; Giles, L. Citeseer x: A scholarly big dataset. In *European Conference on Information Retrieval*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 311–322.
12. Roux, V.; Bril, B.; Karasik, A. Weak ties and expertise: Crossing technological boundaries. *J. Archaeol. Method Theory* **2018**, *25*, 1024–1050.
13. Ghaffar, F.; Buda, T.S.; Assem, H.; Afsharinejad, A.; Hurley, N. A framework for enterprise social network assessment and weak ties recommendation. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, Spain, 28–31 August 2018; pp. 678–685.
14. Wang, X.; Sukthankar, G. Multi-Label Relational Neighbor Classification using Social Context Features. In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 464–472.
15. Perozzi, B.; Al-Rfou, R.; Skiena, S. Deepwalk: Online learning of social representations. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 701–710.
16. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2016**, arXiv:cs.LG/1609.02907.
17. Samek, W.; Wiegand, T.; Müller, K.R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv* **2017**, arXiv:1708.08296.
18. Samek, W. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer Nature: Berlin/Heidelberg, Germany, 2019; Volume 11700.
19. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2017; pp. 4765–4774.
20. Wu, F.; Zhang, T.; Souza, A.H.d., Jr.; Fifty, C.; Yu, T.; Weinberger, K.Q. Simplifying graph convolutional networks. *arXiv* **2019**, arXiv:1902.07153.
21. Zhang, Z.; Cui, P.; Zhu, W. Deep learning on graphs: A survey. *arXiv* **2018**, arXiv:1812.04202.
22. Wang, M.; Yu, L.; Zheng, D.; Gan, Q.; Gai, Y.; Ye, Z.; Li, M.; Zhou, J.; Huang, Q.; Ma, C.; et al. Deep graph library: Towards efficient and scalable deep learning on graphs. *arXiv* **2019**, arXiv:1909.01315.
23. Angulo, M.T.; Lippner, G.; Liu, Y.Y.; Barabási, A.L. Sensitivity of complex networks. *arXiv* **2016**, arXiv:1610.05264.
24. Zhang, J.X.; Chen, D.B.; Dong, Q.; Zhao, Z.D. Identifying a set of influential spreaders in complex networks. *Sci. Rep.* **2016**, *6*, 27823.
25. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2016; pp. 3844–3852.
26. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324.
27. Amin, J.; Sharif, M.; Yasmin, M.; Fernandes, S.L. Big data analysis for brain tumor detection: Deep convolutional neural networks. *Future Gener. Comput. Syst.* **2018**, *87*, 290–297.
28. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.

29. Shuman, D.I.; Narang, S.K.; Frossard, P.; Ortega, A.; Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* **2013**, *30*, 83–98.
30. Sandryhaila, A.; Moura, J.M. Discrete signal processing on graphs. *IEEE Trans. Signal Process.* **2013**, *61*, 1644–1656.
31. Such, F.P.; Sah, S.; Dominguez, M.A.; Pillai, S.; Zhang, C.; Michael, A.; Cahill, N.D.; Ptucha, R. Robust spatial filtering with graph convolutional neural networks. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 884–896.
32. Patacchini, E.; Zenou, Y. The strength of weak ties in crime. *Eur. Econ. Rev.* **2008**, *52*, 209–236.
33. Ruef, M. Strong ties, weak ties and islands: Structural and cultural predictors of organizational innovation. *Ind. Corp. Chang.* **2002**, *11*, 427–449.
34. Montgomery, J.D. Job search and network composition: Implications of the strength-of-weak-ties hypothesis. *Am. Sociol. Rev.* **1992**, *57*, 586–596.
35. Ryan, L. Looking for weak ties: Using a mixed methods approach to capture elusive connections. *Sociol. Rev.* **2016**, *64*, 951–969.
36. Conover, M.D.; Ratkiewicz, J.; Francisco, M.; Gonçalves, B.; Menczer, F.; Flammini, A. Political polarization on twitter. In Proceedings of the International Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.
37. Fiorina, M.P.; Abrams, S.J. Political polarization in the American public. *Annu. Rev. Political Sci.* **2008**, *11*, 563–588.
38. Fan, R.; Xu, K.; Zhao, J. Weak ties strengthen anger contagion in social media. *arXiv* **2020**, arXiv:2005.01924.
39. Sandstrom, G.M.; Dunn, E.W. Social interactions and well-being: The surprising power of weak ties. *Personal. Soc. Psychol. Bull.* **2014**, *40*, 910–922.
40. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43.
41. Grindrod, P.; Parsons, M.C.; Higham, D.J.; Estrada, E. Communicability across evolving networks. *Phys. Rev. E* **2011**, *83*, 046120.