Semi-Local Integration Measure of Node Importance

Ban Kirigin, Tajana; Bujačić Babić, Sanda; Perak, Benedikt

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

Article

Semi-Local Integration Measure of Node Importance

Tajana Ban Kirigin [1,*] , **Sanda Bujačić Babić** [1] and **Benedikt Perak** [2]

1   Department of Mathematics, University of Rijeka, R. Matejčić 2, 51000 Rijeka, Croatia; sbujacic@uniri.hr
2   Faculty of Humanities and Social Sciences, University of Rijeka, Sveučilišna Avenija 4, 51000 Rijeka, Croatia; bperak@uniri.hr
*   Correspondence: bank@uniri.hr

Abstract: Numerous centrality measures have been introduced as tools to determine the importance of nodes in complex networks, reflecting various network properties, including connectivity, survivability, and robustness. In this paper, we introduce Semi-Local Intregation (SLI), a node centrality measure for undirected and weighted graphs that takes into account the coherence of the locally connected subnetwork and evaluates the integration of nodes within their neighbourhood. We illustrate SLI node importance differentiation among nodes in lexical networks and demonstrate its potential in natural language processing (NLP). In the NLP task of sense identification and sense structure analysis, the SLI centrality measure evaluates node integration and provides the necessary local resolution by differentiating the importance of nodes to a greater extent than standard centrality measures. This provides the relevant topological information about different subnetworks based on relatively local information, revealing the more complex sense structure. In addition, we show how the SLI measure can improve the results of sentiment analysis. The SLI measure has the potential to be used in various types of complex networks in different research areas.

Keywords: centrality measure; node importance; complex networks; applications of graph data processing; lexical graph analysis; sentiment analysis

1. Introduction

In the era of Big Data, enormous amounts of data are being collected and analyzed to gain important information and make decisions in a variety of application areas, from managing transportation networks, organizing distribution and delivery, studying biological networks, to organizing the Internet. Graphs ecame the obvious choice for representing the information structure of many data systems. In addition to standard graph theory, a modern wave of mathematical approaches, techniques, and tools are being created, developed, and applied by scientists in various fields, in order to optimize such complex processes.

One of the most important tasks in network analysis is the detection of central or important nodes, which is still a challenge as it depends on the context used. Although many centrality measures are in common use, the category itself is not precisely defined. Many researchers have attempted to provide a mathematical definition of centrality by establishing a set of criteria that measures must meet in order to be considered as centrality measures [1]. Moreover, no formal theory has been developed to explain the differences in behavior among them. According to [1], 'There is certainly no unanimity on exactly what centrality is or on its conceptual foundations, and there is little agreement on the proper procedure for its measurement'. Arguably, that is still the case today.

Among the widely used centrality measures [1–3], some express local node properties while others are more global. The simplest and best known centrality measures, *degree* and *weighted degree* (also called *strength*), reflect strictly local network characteristics by considering only the immediate neighbourhood. Consequently, the (*weighted*) *degree* does not necessarily indicate which node plays the most important role in the whole network. In fact, among the nodes with the same *degree*, some of them may be very central and

the others peripheral, which is not detected by the node *degree* alone. Moreover, bridge nodes are an important class of nodes, because they connect subnetworks of the original network, but may not have the highest *degree*. In contrast, other centrality measures are global [4–6]. For example, *PageRank* [7] and *eigenvector centrality*, take into account information about the nodes and edges of the entire network. For example, *closeness centrality* and *betweenness centrality* involve the shortest paths between the node and all other nodes in the network, and thus reflect global network features. It is important to emphasize that the algorithms for such centrality measures, which require global information to compute the importance of nodes in a network, often have very high time complexity. For this reason, it is of great interest, especially for networks with many nodes and/or intricate structure, to develop new and effective centrality measures that are not computed using the entire network information, but instead focus on smaller subnetworks containing the node [8–12]. This particular task still opens up possibilities for different approaches to the problem, considering various system properties that are analyzed, including efficiency, connectivity, shortest paths, influence, robustness, etc.

Although several approaches to defining and analyzing semi-local node importance have been explored in the literature, focusing on various network properties, including survivability and robustness of the network, to the best of our knowledge none of the existing centralities recognises a high level of node integration. To identify and evaluate the property of interconnectivity among nodes in the neighbourhood of complex networks, in this paper we introduce *Semi-Local Integration* (*SLI*) centrality measure. This centrality evaluates nodes according to how strongly they are integrated in the local subnetwork. This is obtained by taking into account the *weighted degree* centrality of the node itself as well as the *weighted degree* of the nodes in its slightly wider neighbourhood. In addition, the coherence of the neighbouring subnetwork is considered.

We illustrate *SLI* node importance differentiation among nodes in lexical networks and demonstrate its potential in natural language processing (NLP). We implemented the *SLI* centrality measure in the ConGraCNet web application [13], which features the graph-based methodology developed for various lexical tasks. Using the *SLI* measure, we have improved lexical tasks that can take advantage of the *SLI*'s ability to discriminate the local important nodes in a coordination collostruction weighted graph, such as tasks of sense structure and word sentiment analysis. By using relatively local graph information, we were able to obtain a computationally efficient centrality measure and thus obtain relevant information about different subgraphs, and use the obtained *SLI* values of subgraph nodes in propagating subgraph features such as the associated sense and sentiment potential. This implementation shows that the *SLI* centrality is particularly suitable for applications in complex networks and therefore could optimize the analysis of complex network and subnetwork structures, including friend-of-a-friend (FoF)-based networks such as social networks. The applications clearly go beyond linguistics and social sciences, and extend to other research areas. The Python function implementing the *SLI* measure is available at GitHub repository [14].

The paper is organized as follows. We conclude this section with an overview of the related work. We define the *SLI* centrality measure in Section 2 and demonstrate its suitability for lexical applications in Section 3. We conclude with Section 4, where we also propose future research directions.

Related Work

The evaluation method for the importance of a node in a network can be based on a local, semi-local, or global approach. Among the reasons of having so many different approaches to evaluate centralities are the properties of the systems being revealed by the centrality. For example, in a large information network it is very important to backup servers in order not to lose important data, so a redundant design implies that the most important nodes in this case are those that increase the robustness of the network. Connectivity of the network is another important property in many applications, such as transportation,

security, epidemiology, psychology, social studies and others [15–19]. In this context, bridge nodes feature prominently as key nodes in complex networks, even though they often do not have a high *degree*, but connect important subnetworks. Identifying bridge nodes of a network therefore requires a topological analysis of the node neighborhood, such as was done for the definition of the neighborhood-based bridge node centrality tuple in [20].

The semi-local centrality measure of [9] relates the importance of nodes to the survivability and robustness of networks and identifies key nodes for maintaining the underlying function of the network. The failure of key nodes, such as bridge nodes, has a large impact on the network, with either negative (e.g., loss of communication) or positive interpretation (e.g., controlling the spread of viruses, preventing security breaches, etc.). This is orthogonal to our approach, while in [9] occurrence in a cycle reduces the importance due to the available alternative path in the network, occurrence in more cycles in our intended interpretation of high integration indicates better interconnection of the node into its local subnetwork. Furthermore, the networks considered in [9] are unweighted, while we consider weighted and unweighted networks.

There are several approaches to the semi-local node importance in weighted networks [21–24]. A generalization of *degree* and shortest path was introduced in [24], rating a node the more important, the more it acquires strength in the network. An interesting study of node importance related to prominence and control in the fields of transportation, scientific collaboration, and online communication is presented in [22]. It formalizes the tendency of prominent nodes to establish connections among themselves. A series of increasingly selective 'richness-based clubs' of nodes is considered, based on the weight of edges connecting the 'members'. However, even this approach does not provides insight into qualitative differentiation between nodes in terms of local integration.

2. Semi-Local Intregation Centrality

Centrality is one of the fundamental concepts in graph theory and network analysis. Centrality measures attempt to identify the most important nodes in a network and relate the prominence of the nodes in a network numerically.

We define the SLI centrality measure of a node that depends on several features of its neighbouring subnetwork. The SLI measure of a node depends on the *weighted degree* of the node itself and its neighbours. We classify a node as more important if it is adjacent to more nodes with higher *weighted degree*, i.e., we consider 'friends' of the node and their 'friends'.

The measure also considers the number of cycles that include the node. Cycles play an important role in many graph-theory applications, including chemistry, biology, and network analysis. More specifically, cycles in FoF networks are an indicator of the interconnectedness between friends of friends and their friends. FoF subnetworks with few cycles exhibit low coherence, while nodes that are included in many cycles are part of a well-connected subnetwork and are part of a coherent local community. Therefore, such nodes can be considered more important in the context of integration. As an example, consider the small network shown in Figure 1, in particular nodes v_1, v_7 and v_{70}. Nodes v_7 and v_{70} do not appear in any cycle, while node v_1 is a member of four simple cycles. This indicates a much stronger integration of the node v_1 into the newtork G_1. Note that one of the simple cycles that includes v_1 has a length of four, i.e., it extends the integration range in the network from a local, immediate neighbourhood to a broader, semi-local subnetwork.

Finally, the measure also takes into account the edge weight, which is a measure of the relatedness of the endpoint nodes. For example, node v_{70} can be considered more strongly connected to the rest of the network due to its higher weight than node v_{71} with a lower weight.
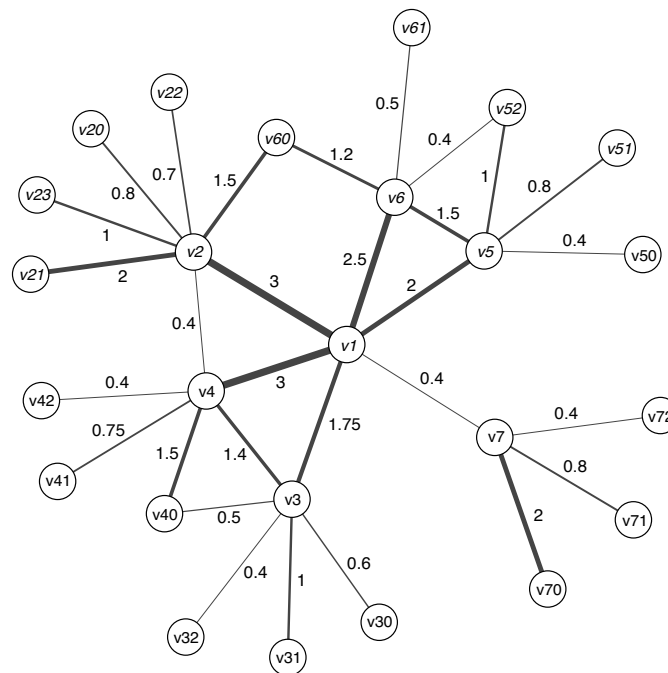
Figure 1. The example network G_1 with edge labels representing weights.

2.1. SLI Definition

In this section, we propose a graph-theoretical definition of node integration. Given an undirected and weighted graph $G = (V, E)$, we use the following notation in the definition of the *SLI* node centrality:

$\deg(v)$—denotes the *degree* of the node $v \in V$;
$d^w(v)$—denotes the *weighted degree* (*strength*) of the node $v \in V$;
$e_{ab} = e_{ba}$—denotes the edge between the nodes $a, b \in V$;
$w(e)$—denotes the weight of the edge $e \in E$;
E_v—denotes the set of edges incident to $v \in V$, $E_v := \{x \mid e_{vx} \in E\}$;
P_G—denotes the cycle basis of the graph G;
$p(e)$—denotes the number of cycles in P_G that contain the edge $e \in E$.

The cycle basis P_G is the minimal set of simple cycles of G that allows every cycle of G to be expressed as a symmetric difference of basis cycles, capturing the local interconnection reach of nodes. The number of local simple cycles that include the edge e, $p(e)$, contributes to the importance of nodes incident to e. We define the *edge cycle factor* of each $e \in E$, $\lambda(e)$, as:

$$\lambda(e) := p(e) + 1. \tag{1}$$

For each node v of G we calculate the node importance by increasing the node's *weighted degree* by the importance contribution of its incident edges:

$$I(a) := d^w(a) + \sum_{e_{ab} \in E_a} I(e_{ab}), \tag{2}$$

where the importance of the edge e, $I(e)$, is defined as follows:

$$I(e_{ab}) := \lambda(e) \cdot \left(d^w(a) + d^w(b) - 2w(e)\right) \cdot \frac{w(e) \cdot d^w(a)}{d^w(a) + d^w(b)} \tag{3}$$

Node importance is then normalized to represent the percentil of node importance share in the graph G. This defines SLI score of the node a:

$$SLI(a) := \frac{I(a)}{S_G} \cdot 100, \tag{4}$$

where S_G is the sum of the unnormalized importance scores of all nodes of G:

$$S_G := \sum_{b \in V} I(b). \tag{5}$$

In a more computational presentation, for a given graph $G = (V, E)$ and the corresponding edge weights and weighted degree scores of its nodes, the calculation of SLI is performed by the following algorithm described in an intuitive metalanguage:

1. Find cycle basis of G, P_G;
2. Find $p(e)$ for all $e \in E$;
3. Find $\lambda(e)$ for all $e \in E$, according to (1);
4. Find the set of edges $E(a)$ for all $a \in V$;
5. Find $I(e)$ for all $e \in E$, according to (2);
6. Find $I(a)$ for all $a \in V$, according to (3);
7. Find $S(G)$, according to (5);
8. Find $SLI(a)$ for all $a \in V$, according to (4).

The Python function implementing the SLI measure is available at GitHub repository [14].

2.2. Discussion on SLI Definition

In a nontrivial connected graph, by definition, none of the three factors in the above definition of edge importance is equal to zero, i.e., each of these three components increases the edge importance. Consequently, each edge incident to a node increases the node's importance. Note that the SLI of isolated nodes is 0, while the importance of edge points of isolated edges is reduced to the weight of the edge.

The SLI scores of all nodes of the graph G_1 shown in Figure 1 are listed in Table 1. Since SLI is a normalized measure, it can help in the interpretation and comparison of importance of nodes in the graph and in the analysis of importance distribution in different graphs. Note that node v_1 has the highest SLI score in G_1 and carries over 40% of the total importance in the graph. Several other top central nodes have high scores, namely $SLI(v_2) = 16.225$, $SLI(v_4) = 13.269$, $SLI(v_6) = 9.614$, $SLI(v_5) = 6.961$ and $SLI(v_3) = 6.664$. Note that among the top ranked nodes, the highest score is more than six times the lowest score, illustrating the non-trivial allocation of importance due to the graph structure. Many nodes in the graph have very low importance, including all leaf-like nodes such as v_{72} with $SLI(v_{72}) = 0.065$. This is an indicator of much lower node integration in the graph.

We now analyze how the definition of the SLI centrality measure reflects the structure of the graph. The *weighted degree* of a node is increased by the importance of its edges. Through the importance of associated edges, the SLI score reflects the integration of the node into the local subgraph.

The first component in (Equation (3)) that contributes to the edge importance is the edge cycle factor $\lambda(e)$, which increases the importance based on the number of simple cycles of G on which the node sits. For example, in the graph shown in Figure 1, the node v_1 is contained in the largest number of simple cycles, four. The nodes v_4 and v_6 are also included in a relatively large number of cycles, namely three, while the leaf-like nodes are obviously not included in any cycle. However, node v_7 is not a leaf-like node, but it is also not included in any cycle. The node v_7 together with the leaves v_{70}, v_{71} and v_{72}, represents a subgraph of G_1 that is not strongly connected to the rest of the graph.

The second component reinforcing the importance of the edge e,

$$d^w(a) + d^w(b) - 2w(e_{ab}),$$

reflects the integration of the endpoints of the edge with the rest of the graph. Note that $d^w(a) - w(e_{ab})$ corresponds to the total contribution of the edges incident to node a, except edge e, to the *weighted degree* of a. This represents the strength of the connection of a with its other neighbours. Thus, this component, denotes the remaining integration of the two endpoint nodes of the edge e_{ab} into the local subgraph. Therefore, the integration of the node itself and the integration of its neighbours into the local subgraph, strengthens the node importance. As an example, consider the edges $e_{v_1 v_7}$ and $e_{v_7 v_{72}}$. Although these edges have the same weight, the endpoint nodes of edge $e_{v_1 v_7}$ are more strongly connected to the rest of the graph. The edge $e_{v_7 v_{72}}$ exhibits a much less overall integration in the graph.

The third component,

$$w(e_{ab}) \cdot \frac{d^w(a)}{d^w(a) + d^w(b)},$$

reflects both the weight of the edge and the (im)balance in the *weighted degree* of its endpoint nodes. The weight of the edge was not reflected in the other components, but should not be disregarded. For example, among the leaf-like nodes connected to node v_2, nodes v_{20}, v_{21}, v_{22} and v_{23}, node v_{21} is incident with the edge with the highest weight. The node v_{21} appears to be the most important among the mentioned nodes, since it is best connected to the rest of the graph through the edge $e_{v_2 v_{20}}$. This can be observed from the SLI scores, since the score $SLI(v_{21})$ is higher than the SLI score of any of the nodes v_{20}, v_{22} and v_{23}. Similarly, the same edge $e_{v_2 v_{20}}$ contributes more to the importance of v_2 than the edges associated with other leaf-like nodes.

The remaining fraction involving the *weighted degrees* of the endpoint nodes a and b adds the larger contribution to the endpoint with the higher *weighted degree*, i.e., the better integrated endpoint. For example, in graph G_1, node v_1 has higher *weighted degree* than node v_2, so edge $e_{v_1 v_2}$ should contribute more to node v_1.

Together with the SLI scores of the nodes of G_1, we list the scores of a number of other centrality measures in Table 1. The comparison shows that the values of SLI behave differently from all the standard scores, namely, *degree*, *weighted degree*, *betweenness* and *PageRank*. Note that the nodes in the table are ordered decreasingly according to SLI, but none of the scores according to another centrality result ordered. Moreover, the importance polarization by SLI is much more pronounced compared to all other centralities, as can be seen in Table 1. Compared, for example, with *betweenness*, SLI reveals some differentiation between leaf-like nodes, albeit to a small extent, e.g., node v_{21} has a higher SLI score than node v_{22} due to the incidence of edge $e_{v_2 v_{21}}$ having a larger weight than edge $e_{v_2 v_{22}}$.

The rather small network example given in Figure 1 already illustrates that overall the SLI displays considerable polarization in ratings of the nodes according to the strength of their interconnection in the local subnetwork.

Table 1. Comparison of different centrality measures of nodes in graph G_1 illustrated in Figure 1.

Node	SLI	Degree	Weigthed Degree	Betweenness	PageRank
v_1	40.225	6	12.65	0.594	0.148
v_2	16.225	7	9.4	0.345	0.118
v_4	13.269	6	7.45	0.246	0.096
v_6	9.614	5	6.1	0.147	0.076
v_5	6.961	5	5.7	0.193	0.077
v_3	6.664	6	5.65	0.259	0.081
v_{60}	1.671	2	2.7	0.024	0.035
v_7	1.238	4	3.6	0.239	0.071
v_{40}	0.969	2	2	0.0	0.029

Table 1. *Cont.*

Node	SLI	Degree	Weigthed Degree	Betweenness	PageRank
v_{21}	0.569	1	2	0.0	0.027
v_{52}	0.546	2	1.4	0.0	0.022
v_{70}	0.39	1	2	0.0	0.039
v_{23}	0.224	1	1	0.0	0.017
v_{31}	0.21	1	1	0.0	0.018
v_{20}	0.166	1	0.8	0.0	0.015
v_{51}	0.16	1	0.8	0.0	0.015
v_{41}	0.15	1	0.75	0.0	0.014
v_{71}	0.149	1	0.8	0.0	0.019
v_{22}	0.139	1	0.7	0.0	0.013
v_{30}	0.11	1	0.6	0.0	0.013
v_{61}	0.088	1	0.5	0.0	0.011
v_{42}	0.067	1	0.4	0.0	0.010
v_{50}	0.067	1	0.4	0.0	0.011
v_{32}	0.067	1	0.4	0.0	0.011
v_{72}	0.065	1	0.4	0.0	0.013

2.3. SLI in Unweighted Graphs

In the case of an unweighted graph G, the SLI measure calculation reduces to the following formula:

$$SLI\,(a) := \frac{I(a)}{S_G} \cdot 100,$$

where the importance of the node a, $I(a)$, is given by:

$$I(a) = \deg(a) + \sum_{e_{ab} \in E_a} I(e_{ab}),$$

and the importance of the edge e, $I(e)$, is:

$$I(e_{ab}) := \lambda(e) \cdot \big(\deg(a) + \deg(b) - 2 \big) \cdot \frac{\deg(a)}{\deg(a) + \deg(b)}.$$

As an example of an unweighted graph and its SLI scores, we take the unweighted version G_2 of the graph G_1 shown in Figure 1. Table 2 shows the corresponding SLI scores, with the nodes ordered as in Table 1, i.e., according to the original, weighted SLI scores. In addition, *degree*, *betweenness* and *PageRank* centralities of G_2 nodes are shown. As with the weighted version of the graph, the SLI ordering of importance does not coincide with any of the other centralities. It exhibits a stronger polarization of importance and a differentiation between the very important nodes and the nodes with the lowest importance.

The normalization of the SLI measure provides a clear interpretation of the relative importance of the nodes in the graph and allows for a comparison between the two versions of the graph, i.e., the comparative analysis of the values shown in Tables 1 and 2.

The role of edge weights is evident in a more pronounced relative scoring and ordering of nodes. Changing just one edge weight, e.g., decreasing the weight of edge $e_{v_1 v_4}$ from 3 to 0.4, has a large effect on node importance ordering, permuting nodes from v_4, v_6, v_3 to v_6, v_3, v_4. The differences in the SLI distribution in the graph can be even more pronounced than in the above example. This is to be expected in complex networks due to the non-uniform distribution of edge weights. Larger differences between edge weights in the network, make the weighted SLI differentiation of node importance even more pronounced.

Table 2. Comparison of different centrality measures of nodes in graph G_2, which is the unweighted version of the graph G_1 illustrated in Figure 1.

Node	SLI	Degree	Betweenness	PageRank
v_1	18.433	6	0.594	0.087
v_2	16.708	7	0.345	0.111
v_4	15.543	6	0.246	0.091
v_6	11.437	5	0.147	0.074
v_3	12.892	6	0.259	0.094
v_5	9.51	5	0.193	0.077
v_7	3.512	4	0.239	0.073
v_{60}	1.943	2	0.024	0.032
v_{40}	1.951	2	0.0	0.032
v_{52}	1.882	2	0.0	0.032
v_{21}	0.427	1	0.0	0.02
v_{23}	0.427	1	0.0	0.02
v_{31}	0.418	1	0.0	0.019
v_{20}	0.427	1	0.0	0.02
v_{41}	0.418	1	0.0	0.019
v_{51}	0.407	1	0.0	0.019
v_{22}	0.427	1	0.0	0.02
v_{71}	0.39	1	0.0	0.022
v_{30}	0.418	1	0.0	0.019
v_{61}	0.407	1	0.0	0.019
v_{70}	0.39	1	0.0	0.022
v_{32}	0.418	1	0.0	0.019
v_{42}	0.418	1	0.0	0.019
v_{50}	0.407	1	0.0	0.019
v_{72}	0.39	1	0.0	0.022

3. Application of *SLI* in Lexical Networks

Graphs are widely used in NLP to represent large amounts of lexical data. Graph-theoretic analysis of lexical networks can reveal features useful for human review and consequently provide insights and ideas for automatic methods (for an overview of graph methods in NLP, see [25]).

In our recent work [26–28], we used graph theory in an interdisciplinary approach to tackle NLP tasks such as semantic similarity identification, sense association and structure, lexical community labeling, and sentiment analysis.

This research has provided the main motivation for defining *SLI* centrality. The node importance expressing high integration of nodes into the local semantic community, was not adequately represented by standard centrality measures. Our goal was therefore to define a centrality measure that reflects the importance of nodes in complex lexical networks and many other networks with the specific structure studied from a similar point of view.

3.1. Application of SLI in the Analysis of Sense Structure

For the underlying graph representation of lexical networks, we extracted lexemes from a tagged corpora based on coordinated syntactic-semantic constructions that reveal a semantic similarity [29,30] . Figure 2 represents such a syntactic-semantic based lexical network with sense associations of a seed word *work*.

Figure 2. ConGraCNet network representation of the sense structure of a seed noun lexeme *work*. The node size reflects the node *SLI* score.

The network is constructed from the 15 most frequent coordination collostructions [31], called friends of the seed lexeme *work*, and the subsequent 15 most important collostructions of each friend lexeme. The lexemes are represented as nodes, while the level of association between two lexemes is expressed by the corpus measure of the collocation, stored in the weight property of the corresponding edge. The weighted undirected FoF network captures the semi-local emergent conceptual relatedness and the semantic domain's structure of a seed lexeme. The subgraph clusters structured by the lexical relations and their prominent nodes reveal the polysemous aspects of the source lexeme. The lexical network clusters represent the cognitively latent associative domains necessary for understanding the overall (poly)semantic structure of a source lexeme. The association strength of a particular node in this network relates to (a) the semantic relatedness to the source lexeme in the overall network (b) the sense contribution of this node to the semantic domain in a sense cluster of a source node. Clearly, more saliently connected nodes in the network carry more relevance for the analysis of lexical sense relatedness and sense potential. This semantic salience of a lexical node can be calculated using the graph centrality measures mentioned above. However, each measure has its own method of calculating the most central nodes, which in this case affects the assignment of a semantically salient nodes. The *SLI* measure captures the semantic salience that is defined by the associative strength (edge weight) of a node, integration with other nodes (degree, weighted degree), and the topological structure of local integration expressed by the graph cycles (betweenness). The integration of all these dimensions gives the *SLI* the necessary local resolution that allows for a more complex discrimination of a node salience than any of the previously mentioned centrality measures.

The centrality scores of the most central nodes of the network shown in Figure 2 are listed in Table 3. The nodes are sorted by *SLI* scores and illustrate the *SLI* importance distribution, which ranges from about 11% to less than 0.2% of relative importance in the entire network. This can be easily seen in Figure 2 by the node size, which reflects the node *SLI* score. The striking polarisation in the scores shows the two most important nodes in the semantic network of the source lexeme, the lexemes *work* and *family*. Then, the group of lexemes that are highly relevant in the associated conceptual domains of the lexeme *work* is revealed. These include the lexemes *time, dedication, research determination, effort, study, commitment, home, project, school, life* and *play*. The remaining nodes of the network are identified as more peripheral in the overall sense of the source lexeme.

Table 3. Centrality in lexical dependency graph of noun lexeme *work* and SenticNet 6 sentiment scores.

Lexeme	Weighted Degree	Betweenness	SLI	SenticNet6 ODV
work	96.35	6219.4667	11.0602	0.9
family	143.21	1615.5262	10.7711	0.883
time	115.68	1057.2845	7.7119	-
dedication	103.43	788.0095	7.3328	0.034
research	117.88	1459.5190	7.0717	0.883
determination	108.08	1037.9310	7.0436	0.231
effort	93.85	982.9595	6.3767	0.037
study	120.33	1687.6012	6.2252	-
commitment	97.82	855.6643	6.1272	0.704
home	107.3	1059.3940	6.0986	-
project	115.6	1705.4262	5.6272	0.9
school	107.21	1129.0250	5.6033	-
life	101.38	1188.8881	4.6721	-
play	96.1	1677	2.5525	-
passion	24.54	0	0.3868	1
business	23.7	21.0357	0.3673	-
money	18.79	0	0.1937	0.065

More importantly, as illustrated in Figure 2 using different node colours, network communities are formed representing different senses of the source lexeme, each containing the nodes with comparable importance in the subnetwork. For example, the source lexeme *work* is a part of community containing the lexemes *dedication, research, commitment*, which have the relatively highest importance in the subnetwork and carry the dominant semantic properties of the lexical sense community. Similarly, the lexemes *time* and *effort* are the most important nodes in the smaller subnetwork that contains other lexemes related to resources, such as *energy, expense, money, marketing, talent, space, date*. These two nodes have relatively close SLI scores that are much higher than those of the other nodes in their subnetwork, and thus carry the core sense of this community. Such diversification of nodes is not achieved by either *weighted degree* or *betweenness*. In particular, *weighted degree* does not emphasize the centrality of the source node, since many nodes in the network have higher *weighted degree* values. On the other hand, the betweenness scores often drop to zero for a significant number of nodes, which then nullifies their participation in the sense analysis of the local community.

The aspect of capturing the fine-grained resolution of semantic relatedness to the source node can be used for pruning the less related nodes. This task is particularly important for larger FoF networks with several hundred nodes from the network. Pruning out less related nodes reduces information noise and semantic drift to less related senses. Furthermore, such a pruned lexical network allows semantic analysis with less data overload, preserving and highlighting the most relevant information. Figure 3 represents a pruned network from Figure 2 with a seed noun lexeme *work*. The network was pruned to the top 50% of nodes according to SLI centrality.

The most important feature for the semantic analysis of such lexical networks is the integration of nodes into subnetworks, which are themselves structured with respect to the source node and form the sense structure of the source concept. The primary meaning of the concept is reflected in the largest community, which usually contains the most important node in the network, the source lexeme node. The size of the preserved communities reflects the sense prevalence or the less frequent word sense/usage.
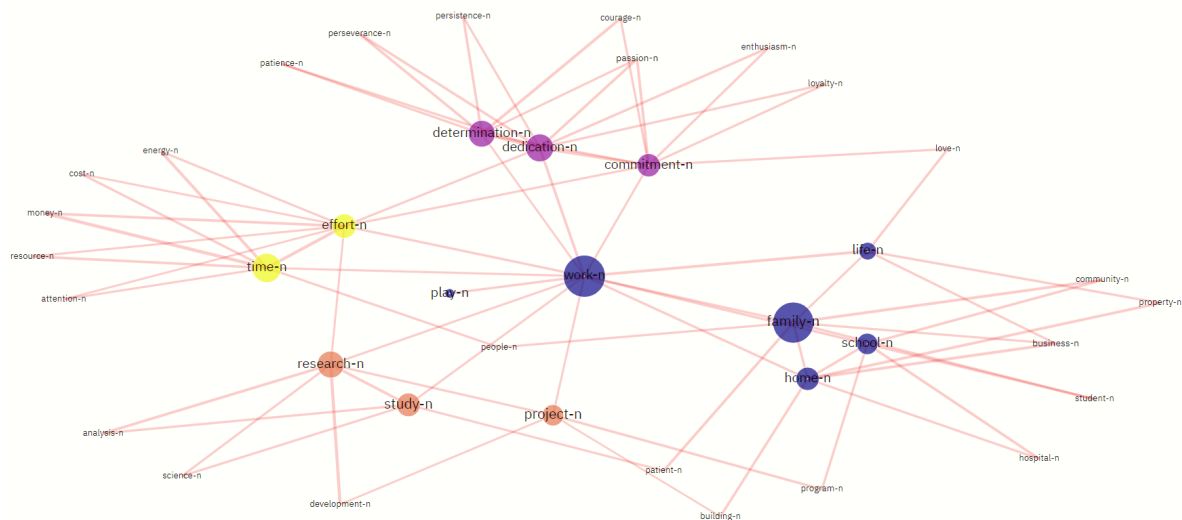
Figure 3. ConGraCNet network representation of the sense structure of a seed noun lexeme *work* pruned to top 50% nodes accoring to *SLI* centrality. The node size reflects the node *SLI* score.

In summary, the centrality obtained with *SLI* has promising applications for lexical network analysis, starting with filtering out less relevant nodes in order to simplify and emphasize the network structure and reduce computational data overload. It also highlights the importance of the source node and enables more qualitative semantic analysis at the level of subnetwork communities identified by a community detection algorithm [32–35] at a selected granularity, i.e., resolution. These properties of the *SLI* measure could advance NLP methods, e.g., improve the results we have obtained in the task of sense labeling [27,28].

3.2. Application of SLI in Sentiment Analysis

Sentiment analysis uses NLP techniques and resources to address affective and subjective phenomena in texts by classifying linguistic expressions from single words (lexemes) to multi-word expressions and longer texts, and assigning them a normalized range of values, typically on a scale from −1 to 1 [36–38].

The simplest linquistic expressions that articulate sentiment are words. The types of words that express sentiment are nouns, adjectives, verbs and adverbs. Some words represent concepts with predominantly culturally associated positive feelings, such as *love, joy, heart*, while others, such as *violence, death, war, failure*, express negative feelings. Classification and numerical evaluation of word sentiment is estimated subjectively by annotators based on psychological evaluation of words or by extending the already annotated dictionaries using various techniques and resources. The information about the sentiment expressed by words is catalogued in sentiment dictionaries, using only positive/neutral/negative classification or a more refined estimates using numerical values and different dimensions.

In the examples we present below, we use the SenticNet 6 sentiment dictionary [36], one of the most comprehensive sentiment dictionaries available. Word sentiment in SenticNet 6 is scored on a number of semantic dimensions, namely the 'polarity_score', 'sensitivity', 'attitude', 'temper', and 'introspection', and is expressed in numerical values between −1 and 1. For example, lexeme *failure* has 'polarity_score' of −0.81 while the following words are scored as follows: *devotion* 1, *heart* 0.9, *love* 0.83, *partner* 0.45, *fighter* 0.343, *flower* 0.054, *roof* 0, *response* −0.2, *desperation* −1.

In our work, on *Sentiment Potential* (SP) [27], we addressed the scarcity of available sentiment dictionaries and problems in sentiment analysis of polysemous and homonymous words, i.e., words with different and potentially unrelated meanings. For example, lexeme

bat denotes both an animal and a baseball device. For another example, lexeme *virus* can be associated with a disease or a computer malware. Moreover, the different senses of a lexeme may have different affective potential. We addressed the above two issues using sentiment propagation, which is intended to reflect both the main affective state and the structure of the feelings expressed by a concept, either a lexeme or a conceptual community of lexemes. The ConGraCNet [13] lexical network, which is a word sense network structure, provides information for an enriched, more complex sentiment representation of sentiment expressed by a word. For an example of SP, see Figure 4 showing SP of the lexeme *work*. As can be seen in Figure 4, the sentiment scores of a single lexeme across its different sense communities are reflected through a vivid illustrative, multi-layered representation.
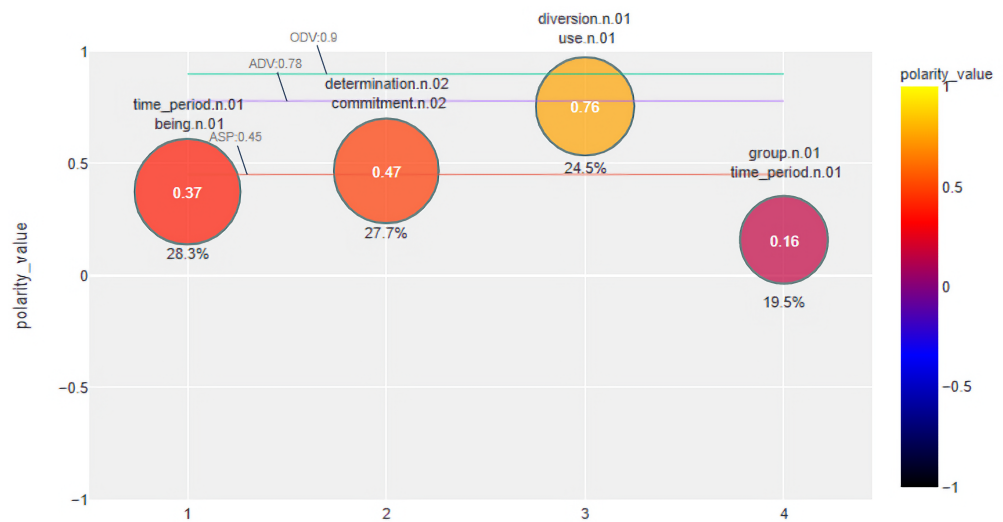
Figure 4. Sentiment potential (SP) of noun lexeme *work* propagated from SenticNet using betweenness centrality.

The SP illustration shown in Figure 4 includes the horizontal line marking the word sentiment score from the SenticNet 6 dictionary, i.e., the *Original Sentiment Value* (*ODV*). Another horizontal line marks the *Assigned Dictionary Value* (*ADV*) [27] of the lexeme which is calculated from the corpus-based coordination dependency lexical graph FoF_a constructed for a chosen source lexeme *a* using the ConGraCNet method [13]. *ADV* of a lexeme *a* is calculated using the available SenticNet 6 *ODV* sentiment values of lexical nodes in the FoF_a graph as follows:

$$ADV(a) := \frac{\sum\limits_{x \in V_a^D} v(x) \cdot b(x)}{\sum\limits_{x \in V_a^D} b(x)}, \tag{6}$$

where $b(x)$ is the *betweenness* measure of the node x in the FoF_a graph, V_a^D is the set of nodes $x \in FoF_a$ with available *ODV* in SenticNet 6, and $v(x)$ is the *ODV* score of lexeme x in SenticNet 6. In addition to assigning the sentiment value to a source lexeme, the same method of propagation can be used to assign the *Graph Sentiment Value* (*GSV*) [27] to a graph. In the case of SP, *GSV* values are assigned to each subgraph of FoF_a representing a sense community in order to estimate the sentiment of the source lexeme in the specific semantic domain.

The third horizontal line in SP illustration marks the mean of the propagated sentiment values over identified lexical subgraphs, Average Sentiment Potential (ASP) [27], represents the average sentiment value of the seed lexeme over its semantic communities:

$$ASP(a) := \sum_{i=1}^{m} GSV(G_a^i) \cdot \frac{\sum_{x \in G_a^i} w(x)}{\sum_{x \in G_a} w(x)}, \tag{7}$$

where m is the number of subgraph communities G_a^i identified in the seed lexeme graph G_a, $GSV(G_a^i)$ is the GSV value of the subgraph G_a^i, and $w(x)$ is the *weighted degree* of the node x in G_a.

The four communities identified in Figure 4 refer to different semantic domains of the lexeme *work*. Each community is represented by a circle whose size represents its share in the total network centrality, which in this case ranges from 28.3% to 19.5%. The vertical placement and the colour of the community representation circle expresses the sentiment value according to the scale shown on the right.

Note that the ODV assigned to the lexeme *work* is very positive, i.e., 0.9. The first approximation obtained using the ConGraCNet method, namely ADV sentiment value of lexeme *work* is lower than the ODV sentiment value, but still rather high, i.e., 0.78. Interestingly, further syntactic-semantic network construction based on the usage of the lexeme in expressed in the corpus and the related analysis, reveals a considerably lower sentiment value related to *work* across all the identified senses.

The centrality measure used in the calculation of SP shown in Figure 4 is betweenness, as originally proposed in [27]. In comparison, SP of the same lexeme computed using SLI scores leads to more differentiated sentiment scores across different sense communities of the lexeme, as shown in Figure 5 and Table 4.

When comparing the two sentiment potentials, various effects of the SLI measure can be noticed when comparing the two sentiment potentials, from the resulting community sizes reflecting the overall importance of the community lexemes, to vertical placements and colours indicating the orientation and intensity of the communities' sentiment score.

The community with lexemes associated around lexemes *work, play, research, study*, for example, displays higher, more positive sentiment value and larger size, i.e., a more prominent, in fact the primary sense of the source lexeme. At the same time, the community containing lexemes *dedication, commitment, determination*, has been reduced in relative size and displays lower sentiment value.
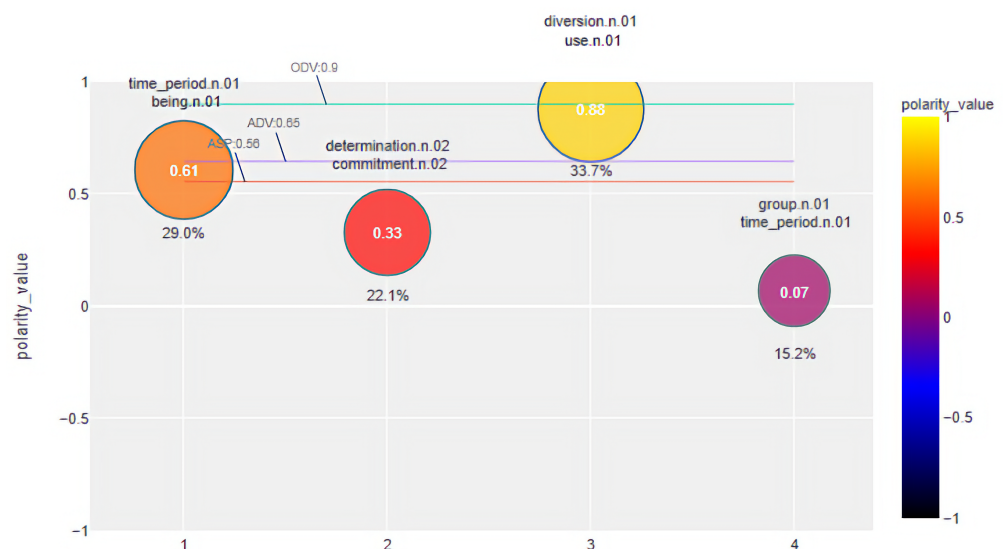

Figure 5. Sentiment potential (SP) of noun lexeme *work* propagated from SenticNet using SLI centrality.

Table 4. Sentiment potential (SP): *GSV* scores of lexical communities of noun lexeme *work* propagated from SenticNet 6.

SenticNet 6	*Work-n*		
Community	**Lexemes**	***Betweenness GSV***	***SLI GSV***
1	*life-n, school-n, home-n, family-n, love-n, property-n, business-n, hospital-n, student-n, community-n, program-n, building-n*	0.37	0.61
2	*dedication-n, commitment-n, determination-n, passion-n, perseverance-n, enthusiasm-n, patience-n, loyalty-n, persistence-n, courage-n*	0.47	0.33
3	*work-n, play-n, research-n, study-n, project-n, patient-n, development-n, analysis-n, science-n*	0.76	0.88
4	*effort-n, time-n, money-n, energy-n, resource-n, cost-n, attention-n, people-n*	0.16	0.07
ODV: 0.9	***betweenness ADV***: 0.78 ***SLI ADV***: 0.65 ***betweenness ASP***: 0.45 ***SLI ASP***: 0.56		

The community sentiment value is propagated from the available original sentiment values of its lexeme nodes. As shown in Table 3, the most important nodes in the third community according to SLI are of positive sentiment score, but of lower values (0.034 and 0.231), while the most positive lexemes from the community, that is nodes *passion* (with the sentiment score of 1) and *commitment* (with the sentiment score of 0.704) are of lower SLI importance. They hence contribute less to the community sentiment score, which is consequently lowered from 0.47 to 0.33.

This effect of accentuating or flattening sentiment differentiation originates in the propagation method of the sentiment value from the available values in the network. Since the community of associated lexemes may contain lexemes with different sentiment values, see Tables 3 and 4, the sentiment value, when propagated as an average of the values in the neighbouring subnetwork, is often diluted or blurred by less important nodes in the community. This issue is mitigated by using a centrality with a high polarization, which is one of the main characteristics of the SLI measure.

3.3. Further Areas of Applications

Besides the linguistic syntactic-semantic applications, that are the main driving force behind the conceptualization of this measure, there are other foreseeable applications of SLI. These include, for example, the analysis of various types of network structures. For instance, SLI can be introduced into graph analysis of vulnerability and identification of bottlenecks in transportation networks [15]. Further applications can be considered for biology structures where a cell, gene, or protein can be considered as a node, and the connecting element as an edge [39], infectious disease spread analysis [40], information network modelling, social networks [41], interaction between social agents such as the influence spreading [42], the impact of cultural norms and legislation [43], etc. The appropriateness of the SLI measure is highlighted in graph analysis of node influence within network structures that involve a weighted type of relation between different entities, components, or agents, with strong topological coherence within a subgraph component.

4. Conclusions

In this paper, we propose a new centrality measure SLI that evaluates the interconnectedness of nodes in a semi-local subgraph. The strong integration of a node within its neighbourhood is reflected by the weighted degree of the node, the weighted degrees of the nodes it is connected to, and the number of cycles it is contained in. The larger all of these factors are, the better the coherence of the neighbouring subnetwork.

The definition of the SLI measure was motivated by applications in lexical networks and has proven useful in the NLP tasks of sense structure representation and sentiment analysis. Further use of the SLI measure appears promising for other NLP tasks, such as sense labeling, where it can help in identifying and classifying hypernym candidates. Moreover, the use of SLI centrality is not limited to lexical networks and opens up studies in various applications of FoF and complex networks in other research areas.

Author Contributions: Conceptualization, S.B.B. and T.B.K.; methodology, T.B.K., S.B.B. and B.P.; software, B.P. and S.B.B.; validation, T.B.K., S.B.B. and B.P.; formal analysis, T.B.K., S.B.B. and B.P.; investigation, T.B.K., S.B.B. and B.P.; resources, B.P.and S.B.B.; data curation, B.P. and S.B.B.; writing—original draft preparation, T.B.K., S.B.B. and B.P.; writing—review and editing, T.B.K., S.B.B. and B.P.; visualization, B.P., S.B.B. and T.B.K.; supervision, B.P., S.B.B. and T.B.K.; project administration, B.P. and T.B.K.; funding acquisition, B.P. and T.B.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The data analyis and representation can be found here: ConGraCNet webapp http://emocnet.uniri.hr/congracnet2/, accessed on 23 January 2022; A list of Sketch Engine corpora and their availability: https://www.sketchengine.eu/corpora-and-languages/corpus-list/, accessed on 23 January 2022; SenticNet datasets can be found at: https://sentic.net/downloads/, accessed on 23 January 2022; WordNet synsets: http://globalwordnet.org/resources/wordnets-in-the-world/, accessed on 23 January 2022 ; WordNet domain and SentiWords 1: https://wndomains.fbk.eu/download.html, accessed on 23 January 2022.

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

References

1. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239.
2. Newman, M. *Networks*; Oxford University Press: Oxford, UK, 2018.
3. Das, K.; Samanta, S.; Pal, M. Study on centrality measures in social networks: A survey. *Soc. Netw. Anal. Min.* **2018**, *8*, 13.
4. Gleich, D.F. PageRank beyond the Web. *SIAM Rev.* **2015**, *57*, 321–363.
5. Li, G.; Li, M.; Wang, J.; Li, Y.; Pan, Y. United neighborhood closeness centrality and orthology for predicting essential proteins. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**, *17*, 1451–1458.
6. Riondato, M.; Upfal, E. Abra: Approximating betweenness centrality in static and dynamic graphs with rademacher averages. *ACM Trans. Knowl. Discov. Data (TKDD)* **2018**, *12*, 1–38.
7. Brin, S.; Page, L. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **1998**, *30*, 107–117.
8. Dong, J.; Ye, F.; Chen, W.; Wu, J. Identifying influential nodes in complex networks via semi-local centrality. In Proceedings of the 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27–30 May 2018; pp. 1–5.
9. Liu, J.; Xiong, Q.; Shi, W.; Shi, X.; Wang, K. Evaluating the importance of nodes in complex networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *452*, 209–219.
10. Wang, J.; Rong, L.; Guo, T. A new measure method of network node importance based on local characteristics. *J. Dalian Univ. Technol.* **2010**, *50*, 822–826.
11. Ren, Z.M.; Shao, F.; Liu, J.G.; Guo, Q.; Wang, B.H. Node importance measurement based on the degree and clustering coefficient information. *Acta Phys. Sin.* **2013**, *62*, 128901.
12. Chen, D.; Lü, L.; Shang, M.S.; Zhang, Y.C.; Zhou, T. Identifying influential nodes in complex networks. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 1777–1787.
13. ConGraCNet Application. Available online: https://github.com/bperak/ConGraCNet (accessed on 23 January 2022).
14. Semi-Local Intregation Measure. https://github.com/sbujacic/SLI-Node-Importance-Measure. (accessed on 26 January 2022).
15. Guze, S. Graph theory approach to the vulnerability of transportation networks. *Algorithms* **2019**, *12*, 270.
16. Canright, G.; Engø-Monsen, K. Roles in networks. *Sci. Comput. Program.* **2004**, *53*, 195–214.
17. Lawyer, G. Understanding the influence of all nodes in a network. *Sci. Rep.* **2015**, *5*, 8665.
18. Jones, P.J.; Ma, R.; McNally, R.J. Bridge centrality: A network approach to understanding comorbidity. *Multivar. Behav. Res.* **2021**, *56*, 353–367.

19. Kitsak, M.; Gallos, L.; Havlin, S.; Liljeros, F.; Muchnik, L.; Eugene Stanley, H.; Makse, H. Identification of influential spreaders in complex networks. *Nat. Phys.* **2010**, *6*, 888–893.
20. Meghanathan, N. Neighborhood-based bridge node centrality tuple for complex network analysis. *Appl. Netw. Sci.* **2021**, *47*, 888–893.
21. Newman, M.E. Analysis of weighted networks. *Phys. Rev. E* **2004**, *70*, 056131.
22. Opsahl, T.; Colizza, V.; Panzarasa, P.; Ramasco, J.J. Prominence and control: The weighted rich-club effect. *Phys. Rev. Lett.* **2008**, *101*, 168702.
23. Segarra, S.; Ribeiro, A. Stability and continuity of centrality measures in weighted graphs. *IEEE Trans. Signal Process.* **2015**, *64*, 543–555.
24. Opsahl, T.; Agneessens, F.; Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Netw.* **2010**, *32*, 245–251.
25. Nastase, V.; Mihalcea, R.; Radev, D.R. A survey of graphs in natural language processing. *Nat. Lang. Eng.* **2015**, *21*, 665–698.
26. Perak, B.; Ban Kirigin, T. Corpus-Based Syntactic-Semantic Graph Analysis: Semantic Domains of the Concept Feeling. *Raspr. Časopis Instituta Za Hrvat. Jez. I Jezikoslovlje* **2020**, *46*, 493–532.
27. Ban Kirigin, T.; Bujačić Babić, S.; Perak, B. Lexical Sense Labeling and Sentiment Potential Analysis Using Corpus-Based Dependency Graph. *Mathematics* **2021**, *9*, 1449.
28. Perak, B.; Ban Kirigin, T. Dependency-based Labeling of Associative Lexical Communities. In Proceedings of the Central European Conference on Information and Intelligent Systems (CECIIS 2021), Varaždin, Croatia, 13–15 October 2021; pp. 34–42.
29. Van Oirsouw, R.R. *The Syntax of Coordination*; Routledge: New York, NY, USA 2019.
30. Langacker, R.W. *Investigations in Cognitive Grammar*; De Gruyter Mouton, 2009. Available online: https://doi.org/10.1515/9783110214369 (accessed on 23 January 2022).
31. Stefanowitsch, A.; Gries, S.T. Collostructions: Investigating the interaction of words and constructions. *Int. J. Corpus Linguist.* **2003**, *8*, 209–243.
32. Raghavan, U.N.; Albert, R.; Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* **2007**, *76*, 036106.
33. Latapy, M. Main-memory triangle computations for very large (sparse (power-law)) graphs. *Theor. Comput. Sci.* **2008**, *407*, 458–473.
34. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008.
35. Traag, V.A.; Waltman, L.; Van Eck, N.J. From Louvain to Leiden: Guaranteeing well-connected communities. *Sci. Rep.* **2019**, *9*, 5233.
36. Cambria, E.; Li, Y.; Xing, F.Z.; Poria, S.; Kwok, K. SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020; pp. 105–114. Available online: https://doi.org/10.1145/3340531.3412003 (accessed on 23 January 2022).
37. Sentic. Available online: https://sentic.net/ (accessed on 23 January 2022).
38. Gatti, L.; Guerini, M.; Turchi, M. SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis. *IEEE Trans. Affect. Comput.* **2015**, *7*, 409–421.
39. Gao, W.; Wu, H.; Siddiqui, M.K.; Baig, A.Q. Study of biological networks using graph theory. *Saudi J. Biol. Sci.* **2018**, *25*, 1212–1219.
40. Venkatraman, Y.; Balas, V.E.; Rad, D.; Narayanaa Y,K. Graph Theory Applications to Comprehend Epidemics Spread of a Disease. *BRAIN. Broad Res. Artif. Intell. Neurosci.* **2021**, *12*, 161–177.
41. Chakraborty, A.; Dutta, T.; Mondal, S.; Nath, A. Application of graph theory in social media. *Int. J. Comput. Sci. Eng.* **2018**, *6*, 722–729.
42. Lehmann, S.; Ahn, Y.Y. *Complex Spreading Phenomena in Social Systems*; Springer: Cham, Switzerland, 2018.
43. Pătruţ, B.; Popa, I.L. Graph Theory Algorithms for Analysing Political Blogs. In *Social Media in Politics*; Springer: Cham, Switzerland, 2014; pp. 49–62.