2010 IEEE International Conference on Data Mining

Node Similarities from Spreading Activation

Kilian Thiel and Michael R. Berthold
Nycomed Chair for Bioinformatics and Information Mining
University of Konstanz, Germany
{*Kilian.Thiel, Michael.Berthold*}*@uni-konstanz.de*

***Abstract*—In this paper we propose two methods to derive two different kinds of node similarities in a network based on their neighborhood. The first similarity measure focuses on the overlap of direct and indirect neighbors. The second similarity compares nodes based on the structure of their - possibly also very distant - neighborhoods. Instead of using standard node measures, both similarities are derived from spreading activation patterns over time. Whereas in the first method the activation patterns are directly compared, in the second method the relative change of activation over time is compared. We apply both methods to a real-world graph dataset and discuss the results.**

***Keywords*-spreading activation, graph analysis, node similarities, node signatures**

I. Introduction

Many datasets consist of units of information as well as the relations between these units, and thus can be represented as networks. The need to analyze and explore these networks in various ways to get new insights of the underlying data is growing just as the amount of data. Experts and analysts do not always know exactly what to look for, or where. Methods that suggest unknown, interesting and potentially relevant pieces of information around a certain topic can help to find a focus, induce new ideas, or support creative thinking. Finding nodes that are structurally similar to a query node, extracting the communities of these nodes, and comparing these subgraphs can lead to valuable and interesting insights.

There are many well-known methods to characterize and compare the nodes of networks such as several centrality indices to quantify which nodes are more central than others, clustering methods to decompose a set of nodes into groups e.g. communities, or role assignments to classify nodes and determine equivalences (see [1]). Thereby nodes can be considered equivalent based on different properties, such as neighborhood identity, neighborhood equivalence, automorphic mapping, equal role assignments to neighboring nodes, and others. Usually role assignments are based on a crisp notion of equivalence, which means that two nodes are either equivalent or not, and thus belong to the same class or not. Networks from real-world data are usually noisy and irregular, which makes finding equivalent nodes unlikely. Thus a relaxed notion of equivalence, in the sense that nodes are defined similarly to a certain extent, is useful for a robust comparison of nodes [2].

We propose two methods to derive two different kinds of node similarities heuristically via spreading activation, which is a well-known method to query graphs [6], [9], [10]. The first method is based on the comparison of activation patterns, yielding a spatial similarity. The second method is based on the comparison of change of activation pattern, yielding a structural similarity. The focus of this paper is the definition and explanation of these two similarities and their demonstration on Schools-Wikipedia (2008/09), a real-world dataset, in order to illustrate their suitability to find structurally similar nodes and extract their close neighborhood.

II. Related Work

Originally, spreading activation was proposed by Quillian and Collins [6], [7] to query information networks. The method facilitates the extraction of subgraphs, nodes and edges relevant to a given query. Initially the query nodes are activated. The activation is than spread iteratively to adjacent nodes until a termination criterion is reached or the process converges. The subset of activated nodes, their level of activation, as well as the induced subgraph compose the final result. Spreading activation has been applied in many fields of research [6]–[10]. Most of these approaches use a set of common constraints [9] to restrict the dynamics of the process. In [11] it is shown that pure (constraint-free) spreading activation with a linear activation function on a connected and not bipartite graph always converges to the principal eigenvector of the adjacency matrix of the graph.

Approaches that are conceptually similar to the comparison of activation pattern of nodes, from spreading activation processes, to derive node similarities are given in [3]–[5]. These approaches are based on an iterative process, like spreading activation, and consider nodes to be more similar the more their direct and indirect neighborhood overlaps. Their aim is to detect communities and dense clusters.

In [4] node similarities are determined based on random walks, which are iterative processes as well. Here only paths of a certain length are considered when computing the similarity. Taking into account all computed iterations, as in [5] may yield to higher accuracy. In [5] all iteration results are accumulated with a decay to decrease the impact of the global node neighborhood. Since the accumulated and normalized activation values are used as similarities

1550-4786/10 $26.00 © 2010 IEEE
DOI 10.1109/ICDM.2010.108

the method may yield asymmetric similarities on directed graphs.

In our approach we compare the activation pattern by means of a well-known similarity measure, the *cosine similarity*, yielding symmetric values also on directed graphs. Additionally we propose a second node similarity derived from the comparison of activation changes in each iteration. Based on this method, nodes are similar if the structure of their neighborhood is similar, although the neighborhood needs not overlap at all. This yields a completely different similarity compared to those mentioned above. Usually the level of activation itself, which is sometimes normalized or accumulated over the iterations, represents the relevancy or similarity of nodes to a given query. We propose the comparison of activation pattern and their changes to determine similarities between nodes of a network which has, to the best of our knowledge, never been introduced before in the context of spreading activation.

III. Spreading Activation

Activation is spread on a graph $G = (V, E, w)$, with V as the set of nodes $V = \{1, \ldots, n\}$, $E \subseteq V \times V$ as the set of edges and $w(u, v)$ as the weight of the edge connecting u and v, with $u, v \in V$, $w(u, v) = 0$ if $(u, v) \notin E$. For an ease of exposition we assume that graph G is undirected, however our results easily generalize to directed graphs. The activation state at a certain time k is represented by *activation vector* and denoted by $\mathbf{a}^{(k)} \in \mathbb{R}^n$ with $\mathbf{a}_v^{(k)}$ as the activation of node $v \in V$. Each state $k > 0$ is obtained from the previous state $\mathbf{a}^{(k-1)}$. The initial state $\mathbf{a}^{(0)}$ defines the activation of nodes representing the query. In each iteration activation is spread to adjacent nodes activating them as well. The process is usually terminated after a certain number of iterations, activated nodes or convergence.

In our approach we use a linear standard scenario described in [11] for which convergence is shown for non-bipartite connected graphs. Given a graph $G = (V, E, w)$ and an activation state $\mathbf{a}^{(k-1)}$ at time $k-1$, the activation of a certain node v at time k is defined by

$$\mathbf{a}_v^{(k)} = \sum_{u \in N(v)} w(u, v) \cdot \mathbf{a}_u^{(k-1)}, \forall v \in V, \tag{1}$$

with $N(v) = \{u : \{u, v\} \in E\}$ as the set of neighbors of v. This process can be described in matrix notation. With $W \in \mathbb{R}^{n \times n}$ as the weight matrix defined by $(W)_{uv} = w(u, v)$ a single iteration can be stated as $\mathbf{a}^{(k)} = W\mathbf{a}^{(k-1)}$ leading to $\mathbf{a}^{(k)} = W^k\mathbf{a}^{(0)}$ (for directed graphs it holds $\mathbf{a}^{(k)} = \left(W^T\right)^k \mathbf{a}^{(0)}$). To prevent the activation values from increasing heavily or vanishing, the activation vector is normalized by its Euclidean length after each iteration

$$\mathbf{a}^{(k)} = \frac{W^k\mathbf{a}^{(0)}}{\left\|W^k\mathbf{a}^{(0)}\right\|_2}\ . \tag{2}$$

Rescaling does not change the direction of the activation vector, so convergence to the principal eigenvector $\mathbf{v}_1$ of W is still ensured since $\lim_{k\to\infty} \mathbf{a}^{(k)} = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|_2}$.

IV. Node Signatures

Based on the convergence behavior of a spreading activation process started from a certain node, a *signature* for this particular node can be determined, which is described in this section. Based on this signature nodes can be represented and compared.

Convergence of the spreading activation process yields to query independent results. No matter from which node(s) spreading processes have been started initially, the activation state becomes equal after a sufficient number of iterations. From iteration to iteration, activation vectors change their directions towards the direction of the principal eigenvector of the weight matrix W. The change of activation vectors namely the *velocity* depends on the node(s) from which it was started. A *velocity vector* at time k of a spreading process started at v is defined as

$$\delta^{(k)}(v) = \begin{cases} \mathbf{0} & \text{, if } k = 0 \\ \mathbf{a}^{(k)}(v) - \mathbf{a}^{(k-1)}(v) & \text{, else} \end{cases}\ , \tag{3}$$

with $\mathbf{0}$ as a vector of all 0 and $\mathbf{a}^{(k)}(v)$ as the activation vector at iteration k of a spreading process started at node v.

The norm of a velocity vector represents the amount of change, the step size of the process. For each node the corresponding step sizes can be determined during the convergence process and represented as a vector, called *signature vector*. In this work we use the l_2 norm as step size. Based on the step sizes of each iteration k up to a maximum number of iterations k_{max}, with $0 \leq k \leq k_{max}$, the signature vector of each node is defined. This vector provides information about the convergence speed of a spreading process, starting from a certain node v and is denoted as

$$\tau_k(v) = \left\|\delta^{(k)}(v)\right\|_2, \tag{4}$$

with $\tau(v) \in \mathbb{R}^{k_{max}}$.

V. Node Similarities

Two kinds of node similarities can be derived based on the comparison of activation and convergence behaviors of spreading activation processes starting from each node. On the one hand nodes can be considered similar if their activation vectors (see Equation 2) are similar. On the other hand nodes can be considered similar if the change of activation from one iteration to another is similar (see Equation 4).

These two kinds of similarities compare nodes based on two different properties, (direct and indirect) neighborhood overlap or neighborhood similarity. A neighborhood overlap between two nodes means that a part of the neighborhood

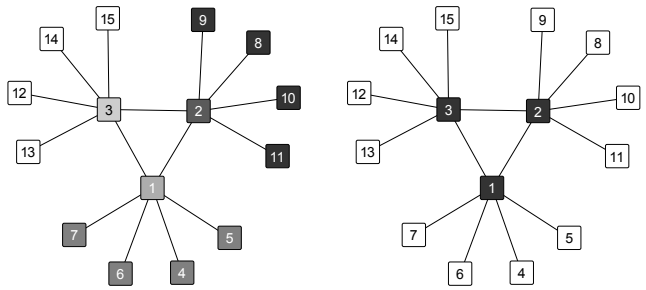


Figure 1. Two node partitionings, indicated by the shading based on two different node properties, equivalent and identical neighborhoods. In 1b the white nodes are structurally equivalent as well as the black nodes, which is determined by comparing the signature vectors. In 1a the leaf nodes are divided into three partitions white, gray, and black, since their particular neighborhood is not identical. Due to the comparison of activation vectors node 3 is more similar to the white nodes, node 2 to the black nodes, and node 1 to the gray nodes than to others.

of these two nodes is identical. This consequently means, the larger the overlap the closer the nodes are in the graph. This property yields a spatial similarity measure and is taken into account when activation vectors are compared. A similar neighborhood of two nodes means that their neighborhood is structurally equivalent to a certain degree but does not necessarily overlap [2]. This can be determined when comparing the change of activation vectors. This property yields a structural similarity measure.

Two node partitionings based on these two different properties are illustrated in Figure 1. The partitioning is indicated by the shading of the nodes. Nodes with the same shade are considered maximally similar (with a similarity value of 1) w.r.t. an equivalent (Figure 1b) or identical (Figure 1a) neighborhood. In Figure 1b the white as well as the black nodes are structurally equivalent since they are automorphic images of each other [1]. In Figure 1a the leaf nodes $\{4, 5, 6, 7\}$, $\{8, 9, 10, 11\}$ and $\{12, 13, 14, 15\}$ are the most similar nodes, due to their identical neighborhood, depicted by the shading gray, black, and white. Even if the leaf nodes are structurally equivalent only those with an identical neighborhood are highly similar. Furthermore the three nodes in the middle $\{1, 2, 3\}$ are not equal based on the comparison of their neighborhood. Node 3 is more similar to $\{12, 13, 14, 15\}$ than to 1 or 2 when comparing their pattern of activation.

The two different similarity measures derived from spreading activation processes allow on the one hand for the identification of structurally similar nodes to a given query node, even if they are located far apart in the graph, and on the other, a densely connected subgraph of direct and indirect neighbors can be extracted for each node. In the following, these two node similarities are formalized and described in detail.

A. Activation Similarity

The first similarity described is based on the comparison of activation vectors and named *activation similarity*. The sequence of activation states of a spreading process started from a certain node describes the node relative to its local and global neighborhood in the graph. Dependent on its neighborhood many or few nodes will be activated and activation will spread quickly or slowly. Nodes close to the initially activated node will become activated sooner than nodes further apart from this node. Furthermore nodes will get activated to a higher level, at least in the primary iterations, if many paths of different lengths exist, connecting them with the initially activated node. Nodes that are similarly connected to a shared neighborhood will induce similar activation states.

The level of activation $\mathbf{a}_i^{(k)}(v)$ of a node $i \in V$ at a time k, induced by a spreading process started at node v, reflects the reachability of i from node v along connecting paths of length k. The more (highly weighted) paths of length k exist connecting i and v, the higher the level of activation. A query node u inducing a similar level of activation $\mathbf{a}_i^{(k)}(u)$ at node i at iteration k is consequently similarly connected to i along (weighted) connecting paths of length k.

Comparing the activation pattern of iterations $k > 1$ allows for the determination of the direct and indirect neighborhood overlap of nodes, whereas measures like the cosine similarity $\sigma_{\cos}(u, v) = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)||N(v)|}}$ or Jaccard index $\sigma_{\text{jaccard}}(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$ based on the characteristic node neighborhood vectors allow for a comparison of the direct node neighborhood only.

In [4] it is stated that in terms of random walks of length k starting from a node v the probability is high for other nodes to be reached if they are located in the same densely connected part or community. For an additional node u, the probability of reaching these nodes is high as well if it is located in the same community. Since random walks are driven by power iterations of the transition matrix of a graph they can be seen as spreading activation processes on a normalized weight matrix.

Considering not only paths of a certain length k as in [4] but all connecting paths of different lengths as in [5] provides a more detailed representation of the local and global neighborhood of a node. Accumulating all activation vectors $\mathbf{a}^{(k)}(v)$ from a spreading process starting from v with a decay α results in a final activation vector $\mathbf{a}^*(v)$ defined by

$$\mathbf{a}^*(v) = \sum_{k=0}^{k_{max}} \alpha^k \mathbf{a}^{(k)}(v), \tag{5}$$

with $0 < \alpha < 1$. The decay α decreases the impact of longer paths and ensures convergence for $k_{max} \to \infty$ for l_2 normalized systems [11]. It is reasonable to decrease the contribution of longer paths to keep more information about

the local neighborhood of v. The above mentioned form is closely related to the centrality index of Katz [12]. We do not want to let the series converge fully since activation vectors of latter iterations do not contribute much to the final activation based on the decay α, and become more and more similar due to convergence of the spreading processes. We chose k_{max} based on the convergence behavior of the underlying graph.

Before a similarity on the final activation vectors is defined we need to take into account that nodes with very high degrees will be activated to a higher level. They are more likely to be reached even if they are not located in the same dense region as the node from which activation has spread initially. To take this into account it is necessary to normalize the final activation by the degree of the corresponding node [4]. The degree normalized final activation vector is thereby denoted as

$$\hat{\mathbf{a}}^*(v) = D^{-\frac{1}{2}}\mathbf{a}^*(v) = D^{-\frac{1}{2}}\left(\sum_{k=0}^{k_{max}} \alpha^k \mathbf{a}^{(k)}(v)\right) \quad (6)$$

with D as the (weighted) degree matrix defined by $(D)_{ii} = d(i), (D)_{ij} = 0$ for $i \neq j, \forall i$ and $d(i) = \sum_{j=1}^{n}(W)_{ij}$.

Based on these normalized final activation vectors we define the activation similarity between two nodes u and v as:

$$\begin{aligned}\sigma_{\text{act}}(u,v) &= \cos(\hat{\mathbf{a}}^*(v), \hat{\mathbf{a}}^*(u)) \\ &= \frac{\sum_{i=1}^{n} \mathbf{a}_i^*(u)\mathbf{a}_i^*(v)d(i)^{-1}}{\|\hat{\mathbf{a}}^*(u)\|\,\|\hat{\mathbf{a}}^*(v)\|}\ .\end{aligned} \quad (7)$$

The more nodes are highly activated in common in both spreading processes, one starting at node u and one at v, the more similar u and v are. Thus this measure allows for a detection of dense communities and requires a direct and indirect neighborhood overlap, as can be seen in Figure 1a. Node 1 is more similar to $\{4, 5, 6, 7\}$ than to 2 or 3 even if 1 is automorphically equivalent to 2 and 3. In [2] this kind of node similarity is categorized as closeness similarity.

The computation of node similarities proposed in [5] can be seen in terms of spreading activation as well. The accumulated and normalized activation values themselves represent the similarities between the activated nodes and the node at which the spreading process started. As stated, their method is applicable only on undirected graphs. For directed graphs the activation values are not necessarily symmetric, yielding asymmetric similarities.

B. Signature Similarity

The second similarity is based on the comparison of the amount of activation changes during spreading activation processes and named *signature similarity*.

Nodes that are similar due to the activation similarity, described in the previous section have to be close to each other in the graph, since the same direct and indirect neighbor nodes need to be activated similarly. The signature similarity is not based on the activation pattern itself but on the amount of change of these patterns. If the structure of the closer neighborhood of two nodes is similar, the change of activation will be similar too, and thus the signature similarity will yield higher values as if the structure is different.

For each node a signature vector can be determined, consisting of velocity vector norms (see Equation 4). The velocity vectors represent the change of direction of the activation vectors and the norms represent the step size between subsequent iterations towards the principal eigenvector. By comparing the signature vectors, a structural similarity can be derived. In this work we use the cosine measure to compare the signature vectors, thus the signature similarity is denoted as:

$$\begin{aligned}\sigma_{\text{sig}}(u,v) &= \cos(\tau(u), \tau(v)) \\ &= \frac{\sum_{k=1}^{k_{max}} \left\|\delta^{(k)}(u)\right\| \left\|\delta^{(k)}(v)\right\|}{\|\tau(u)\|\,\|\tau(v)\|}\ .\end{aligned} \quad (8)$$

A similar step size between two subsequent iterations is yielded from a similar structure, i.e. the nodes $\{1, 2, 3\}$ (black) of Figure 1b are not distinguishable by their signature vectors, due to their structural equivalence. Whereas the activation vectors of these nodes are different, as well as the corresponding velocity vectors, the amount of change of direction of the activation vectors in each iteration is equal. Nodes do not necessarily have to be located in the same densely connected region to have a high similarity. This makes the signature similarity not a closeness but a structural similarity measure. Nodes with a structurally similar neighborhood are similar even if they are located far apart from each other. An overlapping neighborhood is thereby not necessary. This can be seen in Figure 1b, where all the leaf nodes (white) have a signature similarity value of 1, even if their direct neighborhood is not overlapping at all.

VI. Application

To demonstrate our approach we applied the two kinds of node similarities to the Schools-Wikipedia[1] (2008/09) dataset. Our two aims are to:

1) Find result nodes that are structurally similar to a given query node, using the signature similarity.
2) Find nodes that are closely connected (directly or indirectly) to the query node and interesting result nodes, respectively, using the activation similarity, and extract the corresponding subgraphs.

Once structurally similar nodes have been detected and the corresponding communities have been extracted, we manually compare these subgraphs to find structural coherences.

[1] http://schools-wikipedia.org/

A. Schools-Wikipedia

The Schools-Wikipedia (2008/09) dataset consists of a subset of the English Wikipedia[2], with around 5500 articles. Since it is much smaller as the original, it is more suitable for the purpose of evaluation. The selected articles grouped into 154 different categories, consist of 16 main or top level categories, where each article is assigned to at least one category. As in Wikipedia, articles can reference other articles via hyperlinks. In Schools-Wikipedia external links have been filtered.

Based on these hyperlinks we extracted a link graph with articles as nodes and hyperlinks as edges. Additionally we filtered out articles describing years by their important events, since we are interested in persons, events, etc. themselves but not the years in which they occur. The resulting graph is directed and consists of four connected components, whereas three of the components consist only of one node and are filtered as well.

We applied spreading activation processes as described in Section III onto the hyperlink graph, to compute the activation and signature similarities defined in Section V between all nodes. Since the spreading activation processes converge quickly due to the underlying graph structure we only computed the first 10 iterations of each spreading process to compute the similarities. Concerning the activation similarity we used a decay value of $\alpha = 0.3$ to compute the accumulated activation vectors, to focus on the local neighborhood of nodes. The choice of parameters is not discussed in this work. Here it is sufficient to mention that further iterations (> 10) do not contribute significantly to both similarities.

In one of our experiments we wanted to find persons with a similar structure in the network compared to *Linus Torvalds*. Hence we used Linus Torvalds as query node and sorted all articles based on their corresponding signature similarity. Since we focused on structurally similar persons we filtered all articles not belonging to the *People* category. In Table I the 10 most similar nodes of the People category as well as the most dissimilar, based on the signature similarity, compared to Linus Torvalds are listed together with their rank.

It can be seen that Linus himself is the most similar node, which makes sense in terms of the cosine similarity used on activation and signature vectors. Inspecting the structurally similar nodes, the name *Larry Page* attracts our attention as well as *Pope John Paul II*. Larry Page is the fifth structurally most similar node of the People category and Pope John Paul II the most dissimilar.

Next we extracted three subgraphs, containing the 40 nodes most similar to Torvalds, Page, and John Paul II, respectively based on the activation similarity. Figure 2 shows the extracted subgraphs. The layout of all graphs is a centrality layout based on the eigenvector centrality. The higher the eigenvector centrality value of a node, the more central it is positioned. For reasons of visibility and clearness, arrows of directed edges are not drawn. Obtaining an impression of the structure, arrows are not important in this context.

[2]http://en.wikipedia.org/wiki/Main_Page

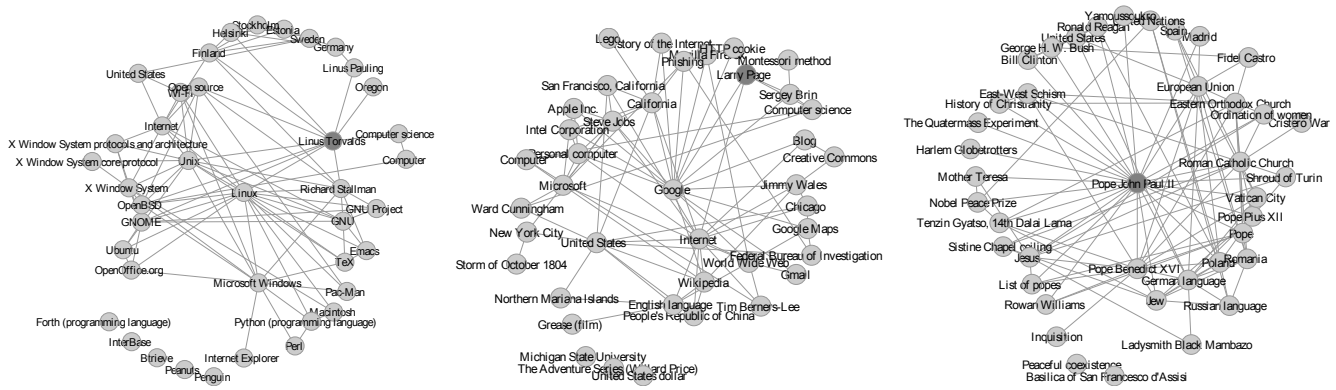
Linus Torvalds	
Rank	Signature similarity
1.	Linus Torvalds
2.	Benjamin Britten
3.	Jackson Pollock
4.	Ward Cunningham
5.	Larry Page
6.	Georgia O'Keeffe
7.	Eilmer of Malmesbury
8.	Emma Roberts
9.	William Renshaw
10.	Marc Pugh
⋮	⋮
695.	Pope John Paul II

Table I
THE 10 STRUCTURALLY MOST SIMILAR NODES, ASSIGNED TO THE PEOPLE CATEGORY, AND THE MOST DISSIMILAR TO LINUS TORVALDS, BASED ON THE SIGNATURE SIMILARITY.

Figure 2a shows the community around Torvalds. Nodes, such as *Linux*, *Open source*, *Helsinki*, etc. are part of his closer community, reasonably, since he studied at the Helsinki University and initiated the development of the open source Linux kernel. The name *Linus Pauling* might at first sight seem exceptional, but on closer inspection it turns out that Torvalds was named after Linus Pauling, the American Nobel Prize-winning chemist. Pauling is a direct neighbor of Torvalds, thus also it make sense for him to be part of Torvalds' closer community.

In Figure 2b the community around Larry Page is illustrated. Apparently Torvalds and Page are not the most central nodes in their communities, but *Linux* or *Google*, respectively. Additionally it can be seen that in both communities two other central nodes exist, *Unix* and *Microsoft Windows* in Figure 2a and *Internet* and *United States* in Figure 2b. The unconnected nodes in both figures have a high activation similarity, otherwise they would not be in the top 40, but have no direct edge to the other connected nodes. When the community is extended to 50 or more nodes, they become connected as well.

Figure 2c shows the community around Pope John Paul II, the structurally most dissimilar node to Torvalds. It can clearly be seen that the connectedness of John Paul II is significantly different to that of Torvalds and Page, since John Paul II is the most central node in his community and almost connected to all other nodes of the extracted subgraph.



(a) (b) (c)

Figure 2. Three subgraphs of the 40 most similar nodes according to Linus Torvalds (Figure 2a), Larry Page (Figure 2b), and Pope John Paul II (Figure 2c) based on activation similarity. The layout of all graphs is a centrality layout, based on the eigenvector centralities of the nodes. Very central nodes, like Linux, Google, or Pope John Paul II are place at a central position.

VII. Conclusion

In this work we have shown how two kinds of similarities to compare nodes in a graph can be derived from spreading activation processes. The activation similarity is based on the comparison of activation vectors and yields a spatial or closeness similarity. The signature similarity is based on the comparison of velocity norms and yields a structural similarity. By applying both kinds of similarities on the one hand enables structurally similar nodes, which are not necessarily located close to each other, to be detected, on the other dense regions around certain nodes, consisting of the closest neighbors, to be extracted and compared. We applied this procedure on the Schools-Wikipedia dataset and preliminary results are very encouraging. The experiments suggested that the combination of these two kinds of similarities could be a promising tool in the area of network exploration.

Acknowledgment

We would like to thank the members of the EU Bison project for many fruitful discussions. The work presented in this paper was supported by a European Commission grant under the 7th Framework Programme FP7-ICT-2007-C FET-Open, project no. BISON-211898.

References

[1] U. Brandes and T. Erlebach, *Network Analysis: Methodological Foundations*, U. Brandes and T. Erlebach, Eds. Springer, 2005.

[2] J. Lerner, "Structural similarity of vertices in networks," Ph.D. dissertation, Universitt Konstanz, Universittsstr. 10, 78457 Konstanz, 2007.

[3] J. Moody, "Peer influence groups: identifying dense clusters in large networks," *Social Networks*, vol. 23, no. 4, pp. 261–283, 2001.

[4] P. Pons and M. Latapy, "Computing communities in large networks using random walks," *Journal of Graph Algorithms and Applications*, vol. 10, no. 2, pp. 191–218, 2006.

[5] E. A. Leicht, P. Holme, and M. E. J. Newman, "Vertex similarity in networks," *Physical Review E*, vol. 73, no. 2, p. 026120, 2006.

[6] M. R. Quillian, "Semantic memory," in *Semantic Information Processing*, M. Minsky, Ed. Cambridge, MA, USA: The MIT Press, 1968, pp. 227–270.

[7] A. M. Collins and E. F. Loftus, "A spreading-activation theory of semantic processing," *Psychological Review*, vol. 82, no. 6, pp. 407–428, 1975.

[8] G. Salton, *Automatic Information Organization and Retrieval*. McGraw Hill, 1968.

[9] F. Crestani, "Application of spreading activation techniques in informationretrieval," *Artif. Intell. Rev.*, vol. 11, no. 6, pp. 453–482, 1997.

[10] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the National Academy of Sciences*, vol. 81, no. 10, pp. 3088–3092, 1984.

[11] M. R. Berthold, U. Brandes, T. Kötter, M. Mader, U. Nagel, and K. Thiel, "Pure spreading activation is pointless," in *Proceedings of the CIKM the 18th Conference on Information and Knowledge Management*, 2009, pp. 1915–1919.

[12] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.