Density based approaches to network analysis

Analysis of Reuters terror news network

Vladimir Batagelj, Andrej Mrvar
University of Ljubljana
vladimir.batagelj@uni-lj.si

Abstract

In the paper some approaches, using program Pajek, to analysis and visualization of networks derived from texts are presented and illustrated on Reuters terror news network. It is shown that using the proposed approaches we can identify in the news network the main themes as we remember them from the news of that period, and also to filter out the secondary themes.

1 Introduction

Centering Resonance Analysis (CRA) is a new text analysis technique developed by Steve Corman and Kevin Dooley at Arizona State University [6, 7]. For demonstration of CRA they produced and analyzed several networks. Among them also the ***Reuters terror news network*** that is based on all stories released during 66 consecutive days by the news agency Reuters concerning the September 11 attack on the U.S., beginning at 9:00 AM EST 9/11/01. The vertices of a network are words (terms); there is an edge between two words iff they appear in the same text unit (sentence). The weight of an edge is its frequency.

In this paper we present some alternative ways to analyze networks on which CRA is based. Instead of centrality (betweeness) based methods used by CRA we explore the edge/vertex density based approaches. We expect that the (relatively) dense subnetworks will reveal the important themes in the news. We first identify the main themes in the total (all 66 days together) network, and afterward visually follow their daily dynamics. Using appropriate normalizations of edges weights we can also identify secondary themes.

All the supporting procedures are implemented in program Pajek [3, 4]. For each approach we provide also a sequence of Pajek's operations to be followed to carry out the analysis and to produce the visualizations. They are given in the Appendix and are indicated by the symbol **A** in the text.

2 News networks

The construction of news networks is a special case of the following scheme, a kind of ***network based data mining***, of analysis of 2-mode networks by transforming them into ordinary (1-mode) networks that are analyzed further using standard network analysis methods.

A 2-***mode network*** is a structure $\mathbf{N} = (U, V, A, w)$, where U and V are disjoint sets of ***vertices***, A is the set of ***arcs*** (directed links) with the initial vertex in the set U and the terminal vertex in the set V, and $w : A \to \mathbb{R}$ is a ***weight***. If no weight is defined we can assume a constant weight $w(u, v) = 1$ for all arcs $(u, v) \in A$. The set A can be viewed also as a relation $A \subseteq U \times V$. A 2-mode network can be formally represented by rectangular matrix $\mathbf{A} = [a_{uv}]_{U \times V}$.

$$a_{uv} = \begin{cases} w_{uv} & (u, v) \in A \\ 0 & \text{otherwise} \end{cases}$$

We denote by $N(u) = \{v \in V : (u, v) \in A\}$ the ***set of neighbors*** of vertex $u \in U$; similary, $N(v) = \{u \in U : (u, v) \in A\}$ is the set of neighbors of vertex $v \in V$.

An approach to analyze a 2-mode network is to transform it into an ordinary (1-mode) network $\mathbf{N}_1 = (U, E_1, w_1)$ or/and $\mathbf{N}_2 = (V, E_2, w_2)$, where E_1 and w_1 are determined by the matrix $\mathbf{A}^{(1)} = \mathbf{A}\mathbf{A}^T$, $a^{(1)}_{uv} = \sum_{z \in V} a_{uz} \cdot a^T_{zv}$ or, considering the network structure,

$$a^{(1)}_{uv} = \begin{cases} \sum_{z \in N(u) \cap N(v)} a_{uz} \cdot a_{vz} & N(u) \cap N(v) \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

Evidently $a^{(1)}_{uv} = a^{(1)}_{vu}$. There is an ***edge*** $(u : v) \in E_1$, $(u : v) = (v : u)$, in $\mathbf{N}_1$ iff $N(u) \cap N(v) \neq \emptyset$. Its weight is $w_1(u : v) = a^{(1)}_{uv}$.

The network $\mathbf{N}_2$ is determined in a similar way by the matrix $\mathbf{A}^{(2)} = \mathbf{A}^T\mathbf{A}$.

It holds: vertices $u, v \in U$ are connected in the network $\mathbf{N}_1$ iff they are weakly connected in the network $\mathbf{N}$. The same holds for network $\mathbf{N}_2$.

From 2-mode networks over large sets U and/or V we can get smaller ordinary networks by first partitioning U and/or V and shrinking the clusters.

The networks $\mathbf{N}_1$ and/or $\mathbf{N}_2$ can be analyzed separately using standard techniques.

There are also different ways to define matrix multiplication – we can select different Abelian semirings. Instead of usual operations $(+, \cdot)$ we can also use (or, and) for $w = 1$, (max, min), (min, $+$) ... In this way we get different ordinary networks.

Also in our case there is a 2-mode network in the background. The construction of CRA networks can be viewed as a transformation of a 2-mode network (units of text, words, contains) into the corresponding 1-mode ***news network*** (words, co-apperance, frequency).

If we ***edge-cut*** a network $\mathbf{N} = (V, E, w)$ at selected level t

$$E' = \{e \in E : w(e) \geq t\}$$

we get a subnetwork $\mathbf{N}(t) = (V(E'), E', w)$, $V(E')$ is the set of all endpoints of the edges from E'. The components of $\mathbf{N}(t)$ – ***islands***, determine different themes. Their number and sizes depend on t. Usually there are many small components. To obtain interesting themes we consider only components of size at least k. The values of thresholds t and k are determined by inspecting the distribution of weights and the distribution of component sizes.

The edge-cut approach is closely related to single-linkage (minimal spanning tree) clustering method. Therefore we can expect the ***chaining effect*** in some results – chaining of themes with common characteristic words.

In some networks we can have also a function $p : V \to \mathbb{R}$ that describes some property of vertices. Its values can be obtained by measuring, or they are computed (for example, centrality indices [9], clustering index, ...).

The ***vertex-cut*** of a network $\mathbf{N} = (V, E, p)$ at selected level t is a network $\mathbf{N}(t) = (V', E(V'), w)$, determined by the set

$$V' = \{v \in V : p(v) \geq t\}$$

and $E(V')$ is the set of edges from E that have both endpoints in V'.

3 Valued cores

We first combined all 66 CRA networks into a single Pajek's temporal network stored on the file `Days.net`. It has $n = 13332$ vertices (different words in the news) and $m = 243447$ edges, 50859 with value larger than 1. There are no loops in the network.

We start our analysis of the ***Terror news*** network by transforming the temporal network into a combined network for all 66 days (union of all time points). First we look for the components in the obtained network. It consists of 22 components – one large of size 13308, 3 of size 2, and 18 isolated vertices. We continue the analysis on the large component, saved as `DaysAll.net`, and the corresponding temporal subnetwork, saved as `DaysCom.net`. See **A1**.

To analyze the combined network we shall use the valued cores [5]. They identify the dense parts of a network.

Let $\mathbf{N} = (V, E, w)$ be a network. For $v \in V$ and $C \subseteq V$ we define the vertex value p

$$p(v; C) = \sum_{u \in N(v) \cap C} w(v, u)$$

where $w(v, u)$ is the frequency of edge (v, u), and $N(v) = \{z \in V : (v : z) \in E\}$ is the neighborhood (set of all neighbors) of $v \in V$.

The t***-core*** of the network is the maximum subset C such that for all $v \in C$ it holds $p(v; C) \geq t$.

In the ***Terror news*** network the weight w is the frequency of co-appearance of given two words (endpoints of the edge). In this case a t-core is the maximum subnetwork in which each its vertex co-appeared in the text at least t times with other vertices from the subnetwork.

There exists a very efficient algorithm to determine t-cores which is also implemented in Pajek. Using it we produce the valued cores partition with step 25. See **A2**. On the basis of distribution of the obtained partition (see Table 1) we decide to look at the cut

Table 1: Valued cores distribution

k	num	interval	k	num	interval
0	0	0 or less	13	36	(300-325]
1	10598	(0- 25]	14	18	(325-350]
2	1081	(25- 50]	15	39	(350-375]
3	479	(50- 75]	16	27	(375-400]
4	300	(75-100]	17	4	(400-425]
5	152	(100-125]	18	14	(425-450]
6	130	(125-150]	19	9	(450-475]
7	159	(150-175]	21	4	(500-525]
8	77	(175-200]	22	2	(525-550]
9	42	(200-225]	26	4	(625-650]
10	56	(225-250]	27	4	(650-675]
11	28	(250-275]	28	2	(675-700]
12	37	(275-300]	35	6	(850-875]

at level $t = 500$ (cluster 20 and higher). At this level there is the last gap $I_{20} = (475, 500]$ in the distribution, and also the number of words in the t-core (= 22) is still small.

Afterward we draw the 500-core, representing the weights by the width of edges. An initial layout is obtained using the Kamada-Kawai procedure, and further improved manually. We export the final picture into SVG (Scalable Vector Graphics) [1, 8]. The obtained SVG picture can be viewed in a web browser and allows the user to interactively select display of cores at different (predefined) levels. See **A3**. The SVG picture is available at:

`http://vlado.fmf.uni-lj.si/pub/networks/Doc/Terror/core500.htm`

From the 500-core (see Figure 1) we can recognize the dominant themes of the news during the 66 days:

- attack on United States, New York and Washington, World Trade Center (twin tower) and Pentagon, using hijacked planes, on Tuesday;
- involvement of United States officials, military and president Bush; United States military strike on Afghanistan;
- involvement of Taliban ruled Afghanistan and Bin Laden;
- war to terrorism.

4 Sequence of pictures for each day

Now, we shall draw a sequence of pictures of daily networks. Since the total (combined) network is large, and also daily networks have several thousands of vertices we have to limit the display to the 'important' part of the network. The importance can be expressed

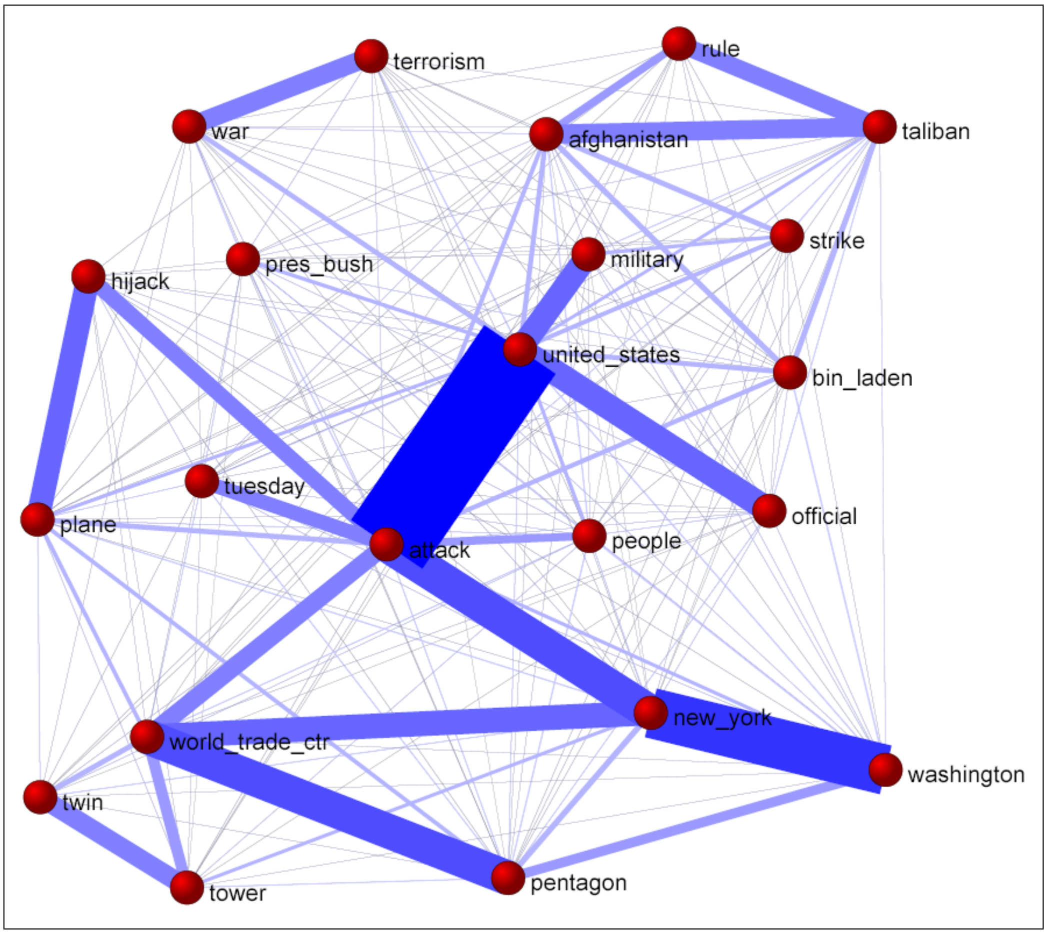

Figure 1: 500-core of the ***Terror news*** network

in many different ways. We shall take for the important part the subnetwork induced by the set W of vertices that are in the `DaysAll` network endpoints of at least one edge with the weight at least t – a vertex-cut with $p(u) = \max_{v \in N(u)} w(v)$. To determine an appropriate value of the threshold t we look at the weights distribution. On the basis of the distribution of the weights we select $t = 50$ and determine the corresponding set W. It has 226 vertices. Then we extract the induced subnetwork and draw it. See **A4**.

We improve the picture manually. In Pajek the coordinates of the vertices of subnetwork are changed in the original network. So, we define with this picture also a layout for the sequence.

We extract from the temporal network `DaysCom` the temporal subnetwork induced by the set W and generate the sequence of 66 pictures. Finally we export the sequence in SVG format. See **A5**.

In Figures 2-5 some snapshots from the obtained SVG sequence
`http://vlado.fmf.uni-lj.si/pub/networks/Doc/Terror/deg50/deg50001.htm`
are presented.

Although the pictures in the sequence are quite different and complicated (see Figure 2) the top levels elements are relatively stable and mainly, no surprise, already known from

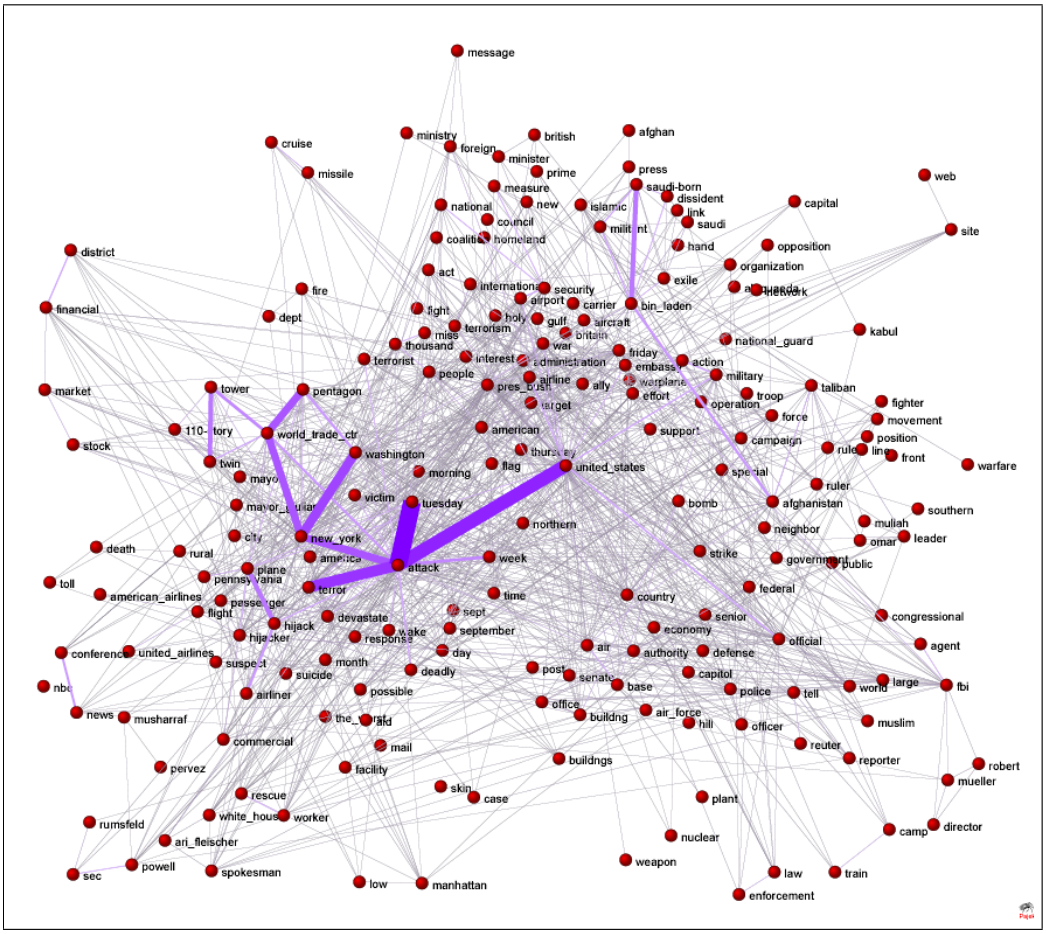

Figure 2: Day 5, 50-vertex-cut $p = \max w$

the main core:

- (suicide) terror attack on United States, New York and Washington, World Trade Center (twin tower) and Pentagon, hijacked plane (airliner), on Tuesday (September), miss people (see Figure 3);
- Taliban (movement, government) ruled Afghanistan; Saudi-born (islamic, militant, fugitive, dissident) Bin Laden (see Figure 4); Al Quaeda (organization, network); (Northern Alliance) opposition.

And a bit less intensive:

- tell Reuter reporter, news conference;
- (national, homeland, airport) security (new measure, council), law enforcement, FBI;
- secretary (Rumsfeld, Powell);
- (district) financial market (stock).

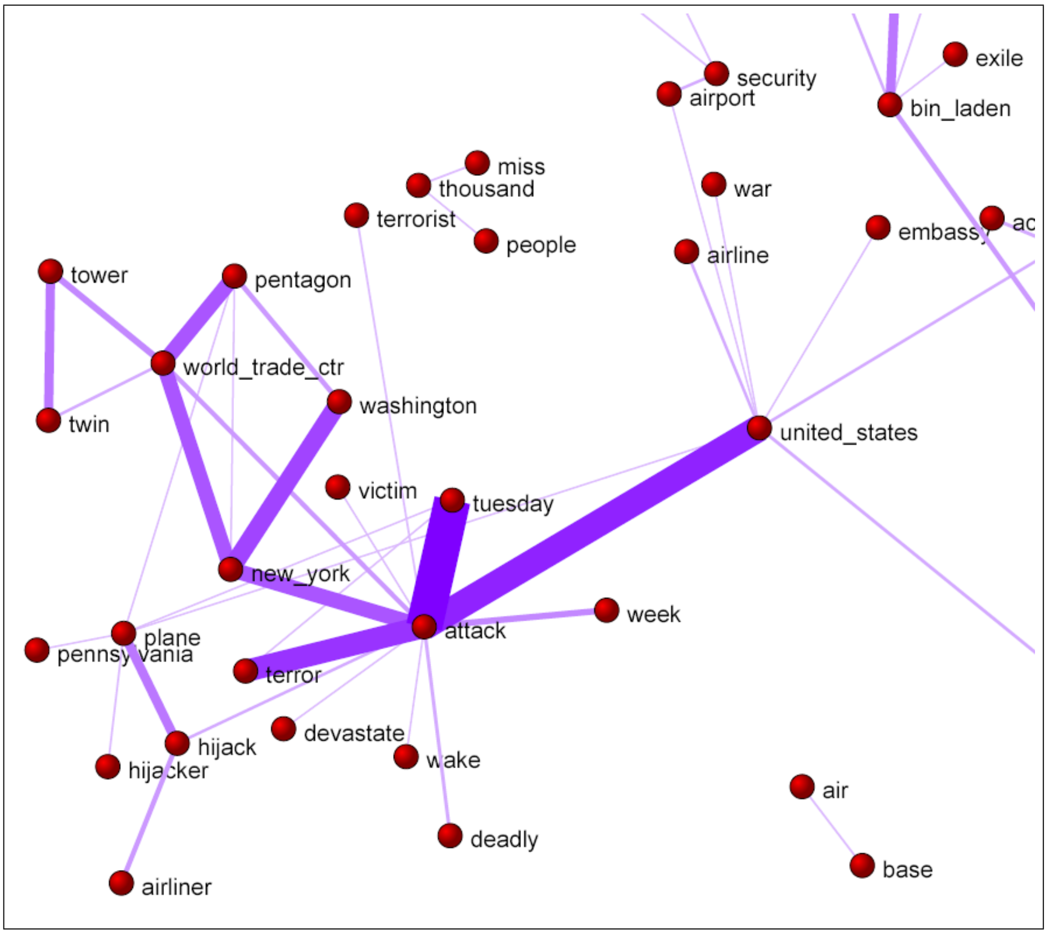

Figure 3: Day 5 zoom: Attack

On the first days the themes death toll (1-3) and rescue worker (1-13) were also strong. Some themes emerged later:

- military action (5); war to terrorism (7);
- anthrax (case, inhale, exposure, bacterium, spore, letter, postal worker) (25, 28-, dominant on 32$^{\text{th}}$ day) (see Figure 5);
- Taliban leader mullah Omar (26); (southern stronghold) Kandahar (27);
- (military, bomb, air) strike on Afghanistan (27);
- (letter) Tom Daschle (office, senate, leader) (36).

Often are present also: Ari Fleischer (white House spokesman), Pervez Musharaf, (American) president Bush, shah Zahir, Robert Mueller (FBI director), Mazar-i-sharif, muslim world, special force, (public) health official, CDC prevention, (nuclear, chemical, biological) weapon.

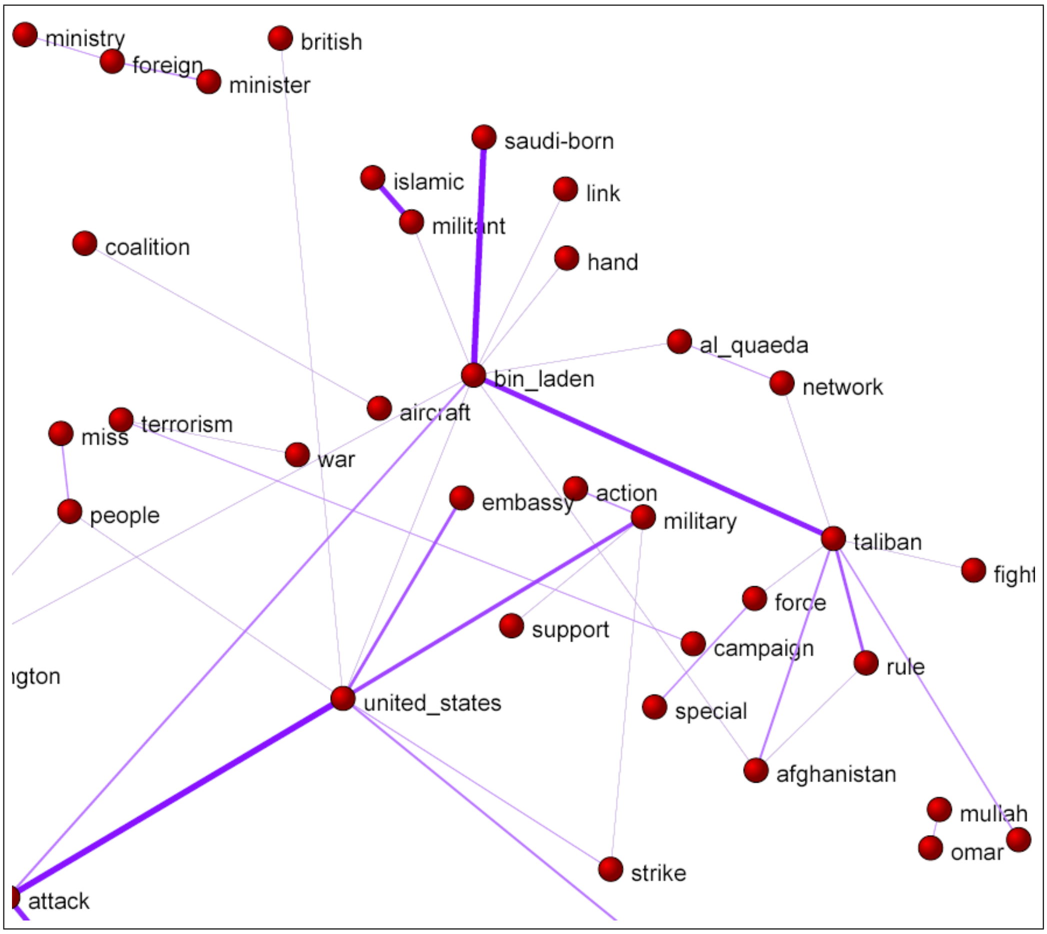

Figure 4: Day 23 zoom: Bin Laden

5 Normalized valued networks

Till now, we were interested in identifying and displaying the most important (dense) parts of the network. But what about the other parts? One approach would be to discard the main part from the network and analyze the residuum.

In the continuation we shall present an alternative approach based on ***compatibility normalization*** of the weights. Because of the huge differences in frequencies of different words it is not possible to compare values on edges according to the raw data. First we have to normalize the network to make the weights comparable. There exist several ways how to do this. Some of them are presented in Table 2.

In the case of networks without loops we define the diagonal weights for undirected networks as the sum of out-diagonal elements in the row (or column)

$$w_{vv} = \sum_u w_{vu}$$

and for directed networks as some mean value of the row and column sum, for example

$$w_{vv} = \frac{1}{2}(\sum_u w_{vu} + \sum_u w_{uv})$$

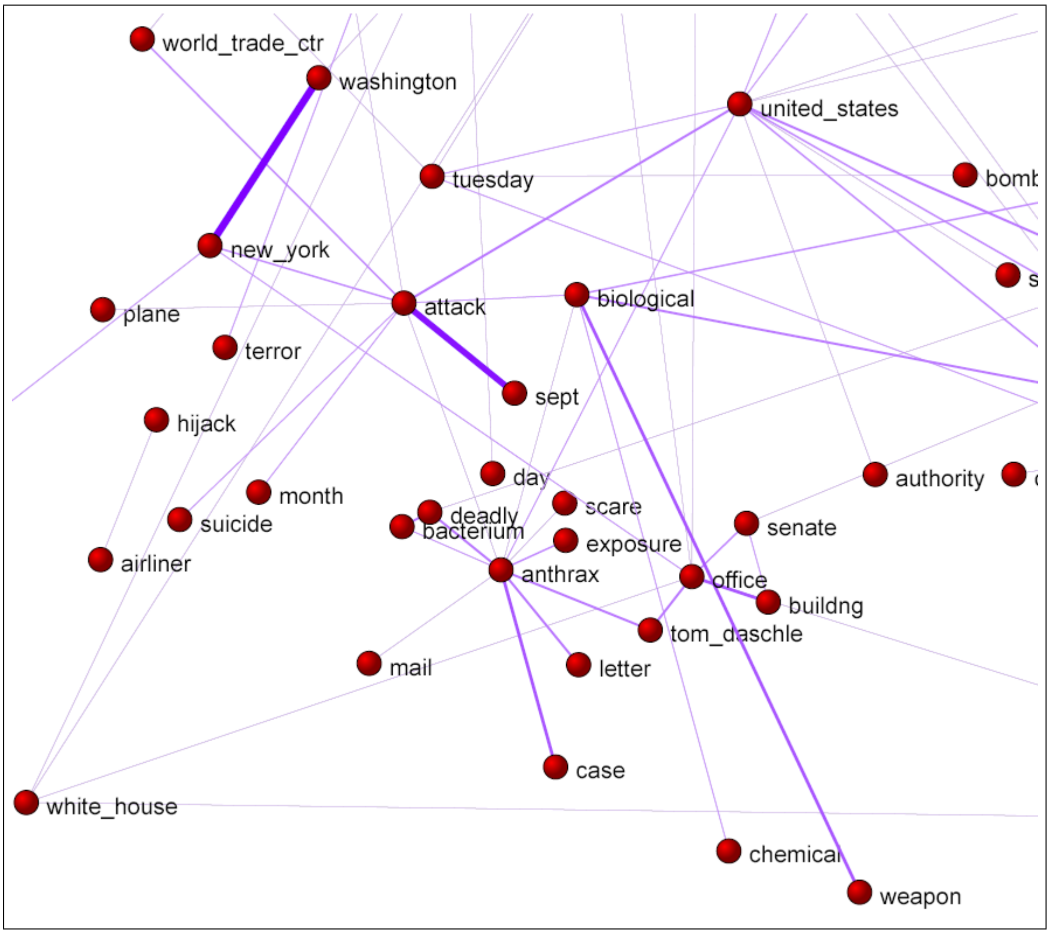

Figure 5: Day 36 zoom: Anthrax

Table 2: Weight normalizations

$$
\begin{aligned}
\mathrm{Geo}_{uv} &= \frac{w_{uv}}{\sqrt{w_{uu}w_{vv}}} & \mathrm{GeoDeg}_{uv} &= \frac{w_{uv}}{\sqrt{\deg\, u \cdot \deg\, v}} \\
\mathrm{Input}_{uv} &= \frac{w_{uv}}{w_{vv}} & \mathrm{Output}_{uv} &= \frac{w_{uv}}{w_{uu}} \\
\mathrm{Min}_{uv} &= \frac{w_{uv}}{\min(w_{uu}, w_{vv})} & \mathrm{Max}_{uv} &= \frac{w_{uv}}{\max(w_{uu}, w_{vv})} \\
\mathrm{MinDir}_{uv} &= \begin{cases} \frac{w_{uv}}{w_{uu}} & w_{uu} \leq w_{vv} \\ 0 & \text{otherwise} \end{cases} & \mathrm{MaxDir}_{uv} &= \begin{cases} \frac{w_{uv}}{w_{vv}} & w_{uu} \leq w_{vv} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

Usually we assume that the network does not contain any isolated vertex.

The normalization approach was developed for quick inspection of (1-mode) networks obtained from 2-mode networks. It was the first time successfully applied in the analysis

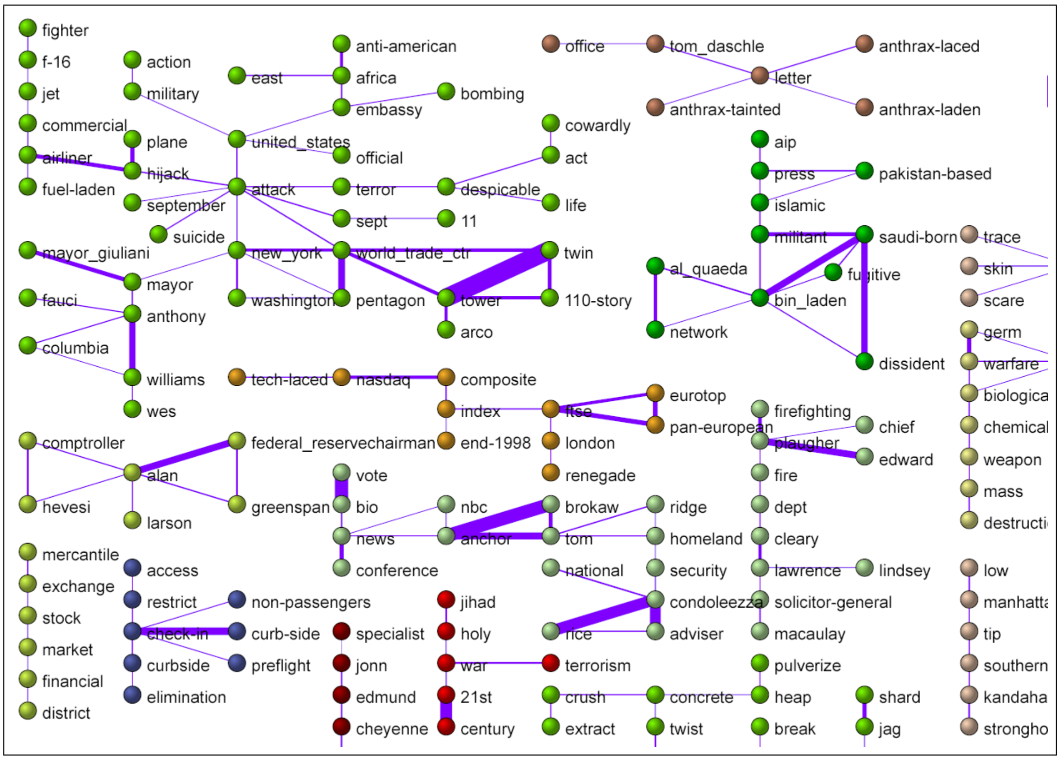

Figure 6: GeoDeg zoom: main themes.

of the 2-mode network (readers, journals, is reading), |readers| > 100000, |journals| = 124 obtained from the readership survey in Slovenia, conducted in 1999 and 2000 by the CATI Center Ljubljana [2, 10].

In this paper we demonstrate the use of GeoDeg, Maxdir and MinDir normalizations.

5.1 GeoDeg normalization

Using the Geo normalization we divide elements of the matrix by geometric mean of both diagonal elements. The standard Geo normalization attains its maximal values on components consisting of a single edge, and is high for strongly 'correlated' vertices – in most of their appearences they appear together. Its application reveals very specific themes. Vertices with large degree have little chance to appear as endpoints of edges with large normalized weight. To give a chance also to these vertices we decided to use the GeoDeg variant of Geo normalization in which the diagonal is filled with degrees of vertices.

The procedure goes as follows. We first determine the degree vector and fill the network diagonal (loops) with it. Then we apply the Geo normalization on this network, and afterward remove the loops. Inspecting the distribution of line values we determine the threshold $t = 0.25$ and cut the network at this level. We are interested only in themes (connected components) of size at least 6. We get a network on 641 vertices with 62 components (themes). We draw them. The color of vertices determines connected component – themes. The final layout

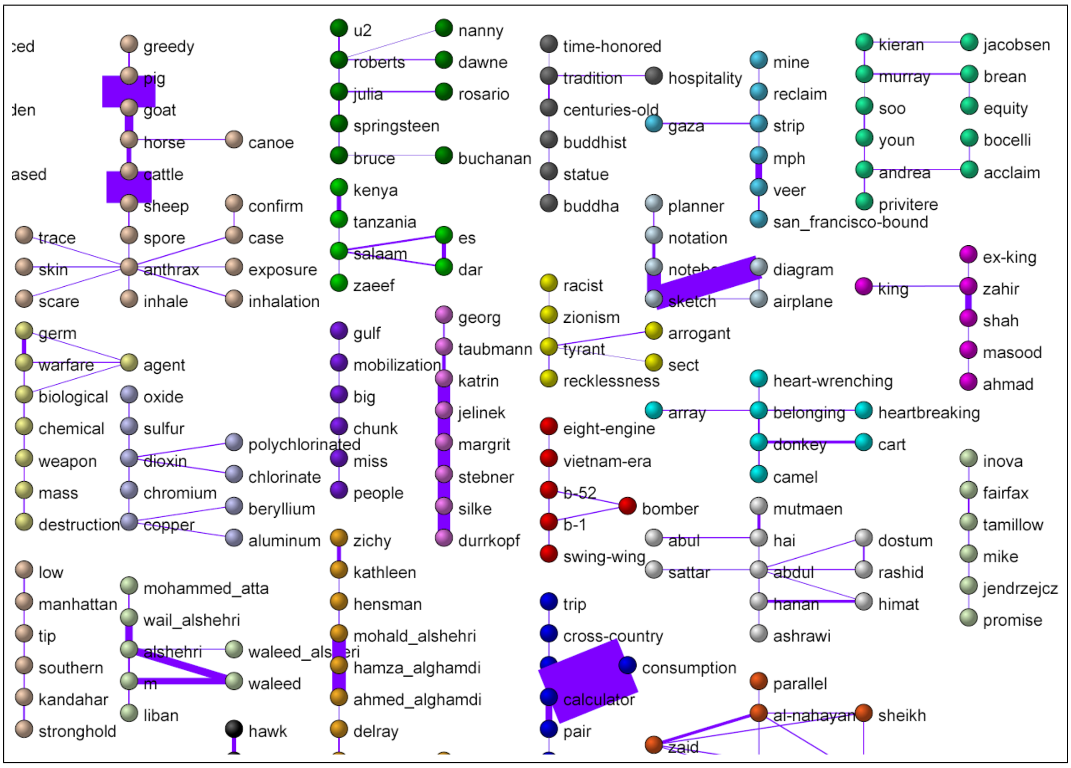

Figure 7: GeoDeg zoom: secondary themes.

`http://vlado.fmf.uni-lj.si/pub/networks/Doc/Terror/GeoDeg.htm` was obtained manually using the Pajek's `grid` option. In Figure 6 and Figure 7 two snapshots of the GeoDeg themes are presented. See **A6**.

In Figure 6 we find essentially the themes (put in broader context) already found in the main core and time sequence analysis: attack, anthrax, Bin Laden, (stock) market and indices, firefighting, security (check-in), jihad, . . .

The themes in Figure 7 are mostly daily news with short duration: sheep-anthrax, Springsteen, Buddha, Bocelli, Margrit, Alshehri, Gaza strip, Zionism, chemistry, calculator, diagram, bomber, donkey, . . .

5.2 MaxDir normalization

The MaxDir normalization transforms an undirected network into a directed one – an arc points from the word with lower frequency to the word with higher frequency; the value on the arc corresponds to the percentage of messages containing the second (terminal) word, that contain also the first (initial) one.

In our case this normalization measures the dependance between words. Large value (close to 1) of the MaxDir weight implies that the two words connected by the arc mainly co-appear.

The procedure to produce and visualize the MaxDir themes is similar to the GeoDeg procedure. The main difference is that for MaxDir normalization we have to compute the row sums vector and add it as a diagonal to the network. After the MaxDir normalization the themes network is cut at level 0.1 and each its component contains at least 6 vertices. We get a network on 502 vertices with 64 components (themes). Also in this case the final layout

`http://vlado.fmf.uni-lj.si/pub/networks/Doc/Terror/MaxDir.htm`

was obtained manually. See **A7**.

In Figure 8 and Figure 9 two snapshots of the MaxDir themes are presented.

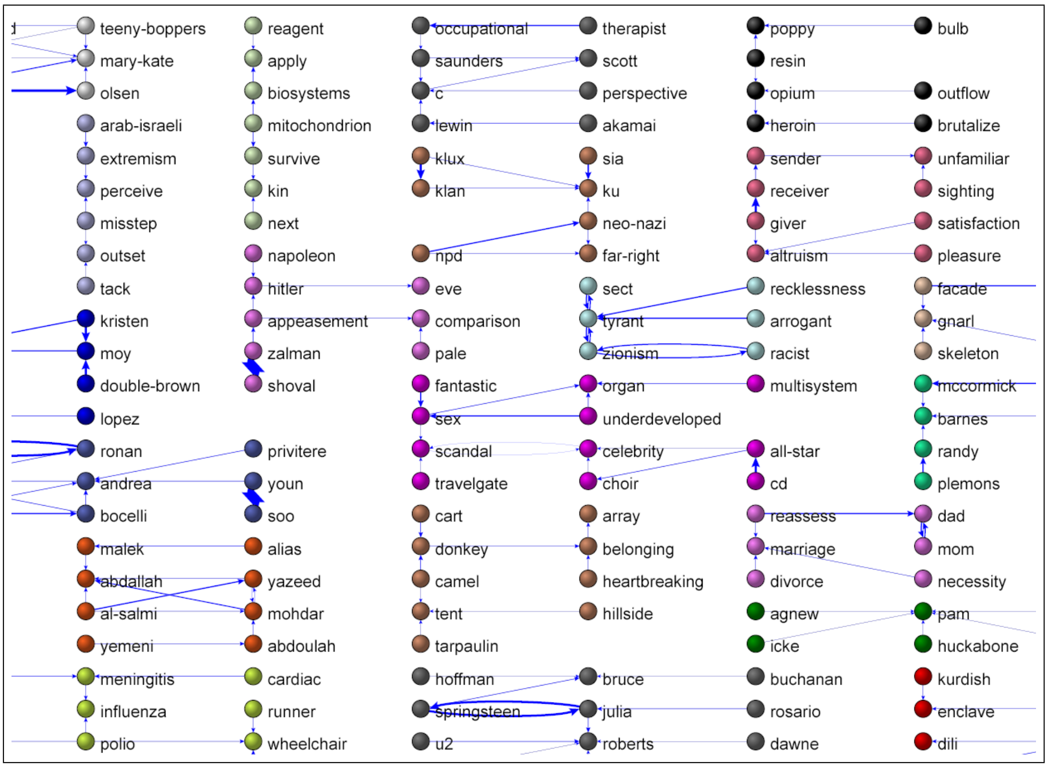

Figure 8: MaxDir zoom

The MaxDir themes are mostly very specialized. For example:

- mom, dad, reassess, marriage, divorce, necessity;
- ku-klux-klan, sia, neo-nazi, npd, far-right;
- reagent, apply, biosystems, mitochondrion, survive, kin, next;
- El-Salvador, Guatemala, Hoduras, Nicaragua, Costa Rica;
- start-up, biotech, genentech, idec, alibek, rituxan, jetblue, low-cost;
- turbo-prop, four-engine, lumber, jstars, ac-130s.

As a by-product of MaxDir computation we can obtain from the row sums vector the list of most frequently used words in the ***Terror news*** network (see Table 3).

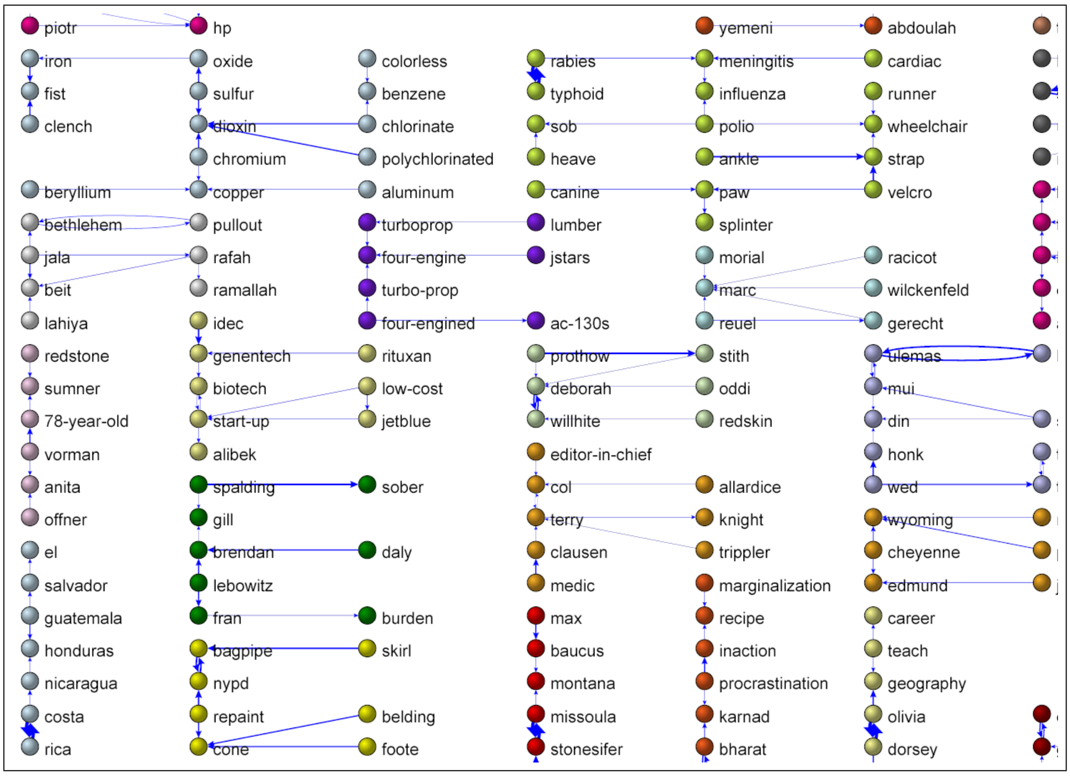

Figure 9: MaxDir zoom

Table 3: The most frequently used words

rank	freq	word	rank	freq	word
1	15000	united states	11	3563	anthrax
2	10348	attack	12	3394	military
3	6266	taliban	13	3078	plane
4	5286	people	14	3006	world trade ctr
5	5176	afghanistan	15	2906	security
6	4885	bin laden	16	2825	american
7	4832	new york	17	2794	country
8	4506	pres bush	18	2689	city
9	4047	washington	19	2679	war
10	3902	official	20	2635	tuesday

5.3 MinDir normalization of valued core

The MinDir normalization also transforms an undirected network into a directed one – an arc points from the less frequent word to the more frequent word. The value on the arc equals to the percentage of text units, that contain the first word, in which also the second word co-appeared.

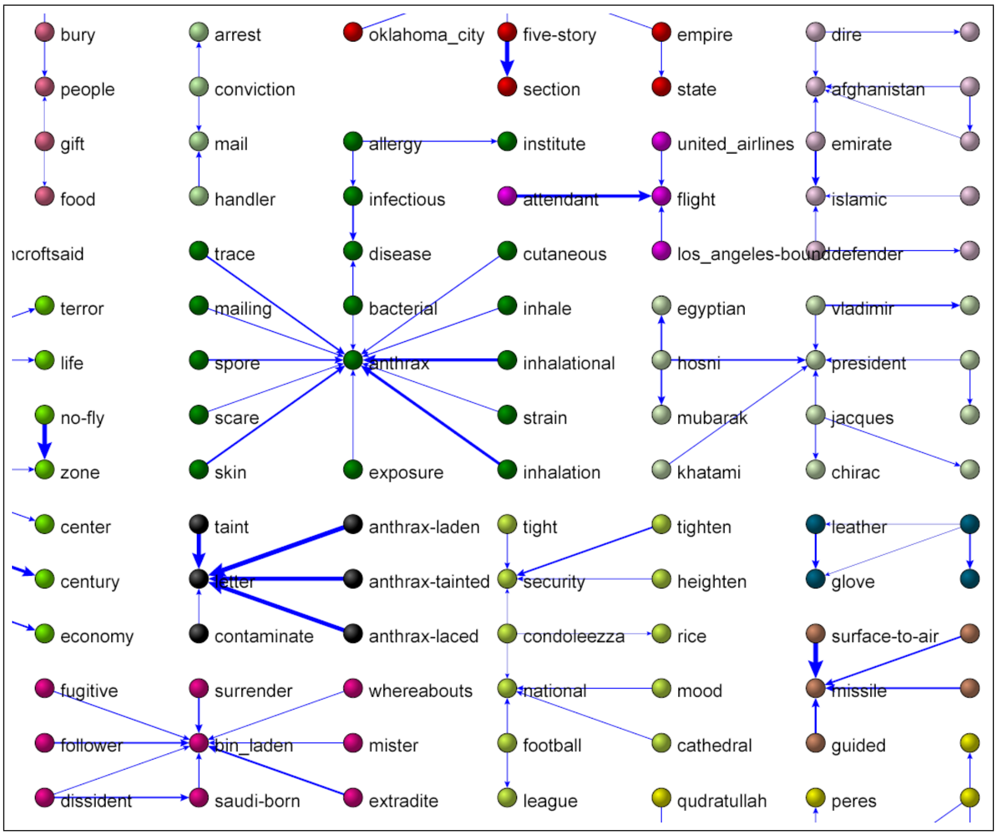

Figure 10: MinDir zoom

The problem with this normalization is that it favorizes vertices with low degree. We can nevtralize this effect by first extracting a valued core from the original network and then applying the MinDir normalization on it. See **A8**.

We extracted from the ***Terror news*** network the 20 core. It has 3122 vertices. After the MinDir normalization we made a cut at level 0.25 and extracted components of size at least 4. The final network

`http://vlado.fmf.uni-lj.si/pub/networks/Doc/Terror/MinDir.htm`

has 312 vertices and 39 weak components. A snapshot from it is presented in Figure 10. As we can see from it, this normalization produces detailed environments of some central words (anthrax, letter, president), but also uncover some secondary themes.

An alternative would be also to use the normalization

$$\mathrm{MinDegDir}_{uv} = \frac{\sqrt{\deg\, u \cdot \deg\, v}}{\Delta}\mathrm{MinDir}_{uv}$$

where $\Delta = \max_{v \in V} \deg\, v$.

6 Conclusions

In the paper we showed that using valued cores and cuts we can identify in the news network the main themes as we remember them from the news of that period. Using different normalizations enables us to filter out also secondary themes. We still have to get the insight on the effects of different normalizations through their applications on real-life networks. Note also, that the Terror news network was built from the Reuters news related to the September 11 attack and therefore it is not the best source for general secondary themes.

All described procedures are supported by program Pajek. The 'granularity' of the obtained results is determined on the basis of distributions. We can use the high level cuts to identify few main themes; but we can also use lower level cuts to reduce the network to smaller interesting parts that are further analyzed using other techniques.

Currently some steps in the analysis (for example: drawing of themes in cuts of normalized networks, detection of changes in daily sequence) still have to be done manually. We need to accumulate additional experiences on different real-life networks to develope the appropriate automatic support for these tasks.

7 Acknowledgments

The paper is an extended version of the presentation at the ***Viszards*** visualization session on the Sunbelt XXII International Sunbelt Social Network Conference, New Orleans, USA, 13-17. February 2002.

This work was partially supported by the Ministry of Education, Science and Sport of Slovenia, Project J1-8532.

A Pajek Details

`Days.net` file

In Figure 11 some segments of `Days.net` are presented from which the structure of Pajek's temporal network description can be seen. The first line says that the network contains 13332 vertices, numbered from 1 to 13332. The line

```
10578 "scope"       [4-7,10,15,17,28,53]
```

tells us that the vertex 10578 represents the word "`scope`" which appeared in the news on days: 4, 5, 6, 7, 10, 15, 17, 28, and 53.

And the line

```
 6844  10575   3 [60]
```

says that on the 60^{th} day the word 6844 "`laboratory`" co-appeared 3 times with the word 10575 "`scientific`".

```
*vertices  13332
 ...
 6842 "label"       [3,31,45,56,57,61]
 6843 "labor"       [2,6,20,25,31,43,53]
 6844 "laboratory"  [2,4,5,15,18,24,28,30,31,37,39-41,45-49,51-53,55,56,60]
 6845 "labour"      [21,22]
 6846 "lace"        [17,59]
 ...
10574 "science"     [2,4,5,14-16,37,40-43,47,53,60]
10575 "scientific"  [15,49,53,60]
10576 "scientist"   [2,6,11,18,21,30,37,41,47,48,55-57,60,63]
10577 "scoff"       [13]
10578 "scope"       [4-7,10,15,17,28,53]
10579 "scorch"      [59]
10580 "score"       [1-3,6,8,9,13,15,26,32,33,35,50,54]
10581 "scoreboard"  [29,50]
 ...
*edges
 ...
 8545   9227   4 [60]
 9227  10935   2 [60]
 1076  11885   2 [60]
 6844  10575   3 [60]
  417   6844   1 [60]
 9288  11741   2 [60]
 ...
```

Figure 11: Pajek's temporal file

A1 Components

```
[read network: Days.net]
Net/Transform/remove/multiple lines/sum values
Net/Components/Weak [1]
Info/Partition [OK][OK]
Operations/Extract from Network/Partition [1][1]
[save network: DaysAll.net ]
[select network: Days]
Operations/Extract from Network/Partition [1][1]
[save network: DaysCom.net ]
```

A2 Valued cores

```
[select network: DaysAll]
Net/Partitions/Valued Core/First Threshold and Step/Input [0][25]
```

A3 Visualization of valued cores

```
Operations/Extract from Network/Partition [20][99]
Draw
Options/Mark Vertices Using/Labels
Options/Lines/Different Widths
Layout/Energy/Kamada-Kawai/Free
[manually improve the layout]
Export/Options [3D Effects on Vertices: check, Gradients: radial]
Export/SVG/Line Values/Options/GreyScale
Export/SVG/Line Values/Options/Different Widths
Export/SVG/Line Values/Nested Classes  [Core500.htm][#10]
```

A4 Layout of the 'important' subnetwork

```
[select network: DaysAll]
Info/Network/Values of lines [#100]
Net/Transform/Remove/lines with value/less than [50]
Net/Components/Weak [2]
Operations/Extract from Network/Partition [1][99]
Draw
Layout/Energy/Fruchterman Reingold/2D
```

A5 Sequence of daily pictures

```
[select network: DaysCom]
Operations/Extract from Network/Partition [1][99]
Net/Transform/Generate in Time/All [1][66][1]
[select network: in time 1]
Draw
[next - previous ...]
[select network: in time 1]
Export/SVG/Current and all Subsequent
Export/SVG/Line Values/Nested Classes [deg50][#15]
```

In the draw window we can inspect the sequence of pictures using `next` / `previous` options. We can also change layout. But attention, the change will effect all the pictures in the sequence.

A6 GeoDeg normalization

```
[select network: DaysAll]
Net/Partitions/Degree/Input
Partition/Make vector
Operations/Vector/Put loops
Net/Transform/2-mode to 1-mode/Normalize/Geo
Net/Transform/Remove/loops
[save network: GeoDeg.net ]
Info/Network/Line values [#100][No]
Net/Transform/Remove/lines with values/lower than [0.25][Yes]
Net/Components/Weak [6]
Operations/Extract from network/Partition  [1] [99]
Net/Components/Weak [1]
Draw/Partition
Layout/Energy/Kamada-Kawai/Free
Move/Grid [50][50]
```

A7 MaxDir normalization

```
[select network   DaysAll.net ]
Vector/Create Identity Vector [13308]
Operations/Vector/Network*Vector [1]
Operations/Vector/Put loops
Net/Transform/2-mode to 1-mode/Normalize/MaxDir
Net/Transform/Remove/Loops
[save network: MaxDir.Net ]
Info/Network/Line values [#100][No]
Net/Transform/Remove/lines with values/lower than [0.1][Yes]
Net/Components/Weak [6]
```

```
Operations/Extract from network/Partition  [1] [99]
Net/Components/Weak [1]
Draw/Partition
Layout/Energy/Kamada-Kawai/Free
Move/Grid [50][50]
```

A8 MinDir normalization

```
[select network   DaysAll.net ]
Net/Partitions/Valued Core/First Threshold and Step/Input [0][20]
Operations/Extract from network/Partition  [2] [999]
Vector/Create Identity Vector [3122]
Operations/Vector/Network*Vector [1]
Operations/Vector/Put loops
Net/Transform/2-mode to 1-mode/Normalize/MinDir
Net/Transform/Remove/Loops
[save network: MinDir.Net ]
Info/Network/Line values [#100][No]
Net/Transform/Remove/lines with values/lower than [0.25][Yes]
Net/Components/Weak [4]
Operations/Extract from network/Partition  [1] [999]
Net/Components/Weak [1]
Draw/Partition
Layout/Energy/Kamada-Kawai/Free
Move/Grid [50][50]
```

References

[1] Adobe SVG viewer. Download Adobe SVG Viewer http://www.adobe.com/svg/

[2] Batagelj V.: Pajek – program for large networks analysis and visualization. Dagstuhl seminar Link Analysis and Visualization, Dagstuhl, July 1-6. 2001 http://vlado.fmf.uni-lj.si/pub/networks/doc/dagstuhl.01/

[3] Batagelj V., Mrvar A.: Pajek. http://vlado.fmf.uni-lj.si/pub/networks/pajek/

[4] Batagelj V., Mrvar A.: Pajek – analysis and visualization of large networks: To be published in *Graph Drawing Software* book, edited by M. Jünger and P. Mutzel, in the Springer series *Mathematics and Visualization*. http://www.ijp.si/ftp/pub/preprints/ps/2003/pp871.pdf.

[5] Batagelj V., Zaveršnik M.: Generalized Cores. http://arxiv.org/abs/cs.DS/0202039

[6] Corman S.R., Kuhn T., McPhee R.D., Dooley K.J.: Studying Complex Discursive Systems: Centering Resonance Analysis of Communication. To appear in Human Communication Research (April 2002, vol. 28 #2)

[7] LOCKS: CRA Analyses of News Stories on the Terrorist Attack. Arizona State University. http://locks.asu.edu/terror/

[8] W3C SVG page. http://www.w3.org/Graphics/SVG

[9] Wasserman S., Faust K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

[10] Zaveršnik M., Batagelj V., Mrvar A.: Analysis and visualization of 2-mode networks. Sixth Austrian, Hungarian, Italian and Slovenian Meeting of Young Statisticians, Ossiach, Austria, October 5-7, 2001. http://www-stat.uni-klu.ac.at/Tagungen/Ossiach/Zaversnik.pdf