P

Pajek and PajekXXL

Vladimir Batagelj[1] and Andrej Mrvar[2]
[1]Faculty of Mathematics and Physics, Department of Mathematics, University of Ljubljana, Ljubljana, Slovenia
[2]Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

Synonyms

Bibliometrics; Graph theory; Kinship analysis; Large networks; Social network analysis; Visualization

Glossary

BOM (byte order mark)	A Unicode character used to signal the byte order of a text file or stream
CPM	Critical path method
GEDCOM (GEnealogical Data COMmunications)	A genealogical software interchange format
GUI	Graphical user interface
MDS	Multidimensional scaling
OR	Operations research
SN5	Network data set on social networks prepared for Viszards session at Sunbelt XXVIII, 2008
SNA	Social network analysis
STRAN	STRuctural ANalysis
SVG (Scalable Vector Graphics)	A WWW picture format
Unicode	The name of the international character set
VOS	Visualization of similarities
VR	Virtual reality

Tool's ID Card

- Tool name, title: **Pajek** and **PajekXXL**, program for analysis and visualization of large networks
- Creation year: November 1996
- Authors: Vladimir Batagelj and Andrej Mrvar
- Range: general network problems with emphasis on large networks
- Copyright: free for noncommercial use
- Type: program
- Scalability: **Pajek**, one billion vertices; **PajekXXL**, two billion vertices
- Platforms: Windows, using emulators runs also on Linux and Mac
- Programming language: Delphi pascal

© Springer Science+Business Media LLC 2018
R. Alhajj, J. Rokne (eds.), *Encyclopedia of Social Network Analysis and Mining*,
https://doi.org/10.1007/978-1-4614-7163-9_310-1

- Orientation: social network analysis, bibliometrics, analysis of genealogies, OR, bioinformatics

Introduction

In the first half of the 1990s, the multimedia computers connected through the Internet provided a new framework for data analysis. They enabled interactive visualization of data and large data sets started to move in the focus of research in data analysis. **Pajek** and **PajekXXL** are programs for analysis and visualization of large networks developed to provide tools for dealing with large networks.

Key Points

Most real-life large networks are sparse – the number of vertices (nodes) and lines (links) are of the same order. This property is also known as a Dunbar number (Hill and Dunbar 2002). The basic idea is that if each vertex has to spend for each link certain amount of "energy" to maintain the links to selected other vertices then, since it has a limited "energy" at its disposal, the number of links should be limited. In human networks, the Dunbar number is between 100 and 150.

From algorithms' complexity theory (Cormen et al. 1990; Batagelj 2009a), it follows that only algorithms of subquadratic complexity are fast enough to be used for analysis of large networks with hundreds of thousands or millions of vertices. It turns out that for some problems for which the general algorithms are too slow, there exist the corresponding algorithms for sparse networks that are subquadratic. Development and implementation of subquadratic algorithms was one of the main goals in the development of **Pajek**.

From the Roman times, the standard approach to deal with something large is the "divide and conquer." In network analysis, this means to support abstraction by (recursive) decomposition of a large network into several smaller subnetworks that can be treated further using more sophisticated methods and control over their interlinks. **PajekXXL** provides tools for identifying important parts of large networks and extract them for further analysis with **Pajek** or using other network analysis tools. **Pajek** also provides the user with some powerful visualization tools for automatic network drawing and for manual improvements of the obtained pictures.

Historical Background and Development

Pajek is a Slovenian word for spider. In Slovenian mathematics the graph theory is one of its strongest fields (Pisanski, Mohar, Klavžar, Marušič, Batagelj, and others). In the 1980s, VB (Vladimir Batagelj) had a series of research projects on graph theory algorithms (Batagelj 1987). A library **graph** for data structure graph was implemented in pascal. In 1990/1991, Anuška Ferligoj and VB were visiting Patrick Doreian at the University of Pittsburgh. They started to develop an optimization approach to blockmodeling (Batagelj et al. 1992a, b; Doreian et al. 2000, 2004) that was later generalized to types of links between clusters (Batagelj 1997) to two-mode networks (Doreian et al. 2004) and by Patrick Doreian and Andrej Mrvar also to signed networks (Doreian and Mrvar 1996, 2009, 2014, 2015, 2016; Mrvar and Doreian 2009; Brusco et al. 2011; Doreian et al. 2013). In Pittsburgh VB developed in pascal a package STRAN for blockmodeling, a very fast algorithm for computing Hummon-Doreian weights in acyclic networks (Hummon and Doreian 1989; Batagelj 2003) and a semiring-based algorithm for computing betweenness centralities (Batagelj 1994).

Andrej Mrvar (AM) was a computer science student at the University of Ljubljana. VB noticed him while he was presenting his home project on Saaty's AH procedure. He became AM's supervisor for diploma, master, and PhD theses. For his master thesis, AM implemented in pascal different graph visualization algorithms. Afterward he started to work on analysis of large networks, which was an emergent topic at that time, for his PhD.

To support AM's work on his PhD, AM and VB in November 1996 decided to collect all their already developed graph analysis programs and to combine them into a single program – **Pajek**. They also decided to make it free for non-commercial use. The first version of **Pajek** was presented on January 29, 1997, on the Wednesday seminar at the Faculty of Mathematics and Physics, University of Ljubljana, and on XVII Sunbelt Conference in San Diego, USA, February 13–16, 1997 (Batagelj and Mrvar 1998). Among the first users of **Pajek** were Wouter de Nooy and Doug White. Since **Pajek** had no user-friendly documentation, VB and AM decided with Wouter to write a book *Exploratory Social Network Analysis with Pajek*. It was published in 2005 and the revised and expanded edition in 2011 by CUP (De Nooy et al. 2011). The book was also published in Japanese in 2009 and in Chinese in 2012. In collaboration with Doug White and later with Klaus Hamberger, different tools for analysis of genealogical data were added to **Pajek** (White et al. 1999; Batagelj and Mrvar 2008). For example, **Pajek** can read genealogies from GED files and supports (bipartite) p-graph representation of genealogies.

In the 1990s, VB and AM contributed several solutions to yearly graph drawing contests (see Fig. 1). They won many prizes but also included in **Pajek** some new tools to address the contests' problems. In Batagelj et al. (1999), they argued that for dense networks, the matrix representation is more readable than the standard graphical representation. They also noticed that Seidman's (1983) notion of cores is the only known formalization of dense parts of the network that can be efficiently determined. The chemists from MATH/CHEM/COMP conference were interested in identification of selected substructures in large organic molecules. For this purpose, the fragments (later reinvented and called motifs by physicists) searching procedure was added to **Pajek** in 1998.

VB was a PhD supervisor also to Matjaž Zaveršnik who studied methods for network decomposition. They developed algorithms for generalized cores (Batagelj and Zaveršnik 2011), islands (Zaveršnik and Batagelj 2004), and short cycle connectivity (Batagelj and Zaveršnik 2007). These methods were also included in **Pajek**.

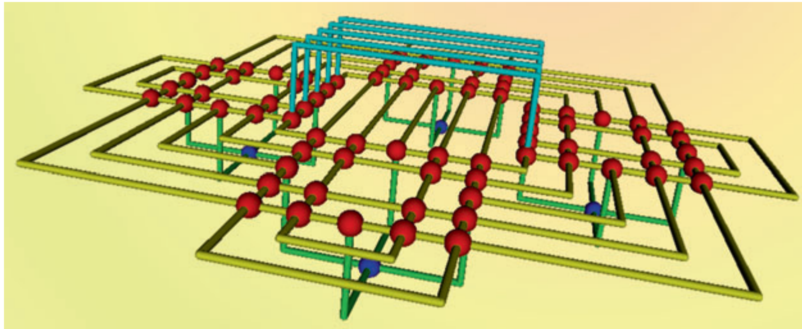
Multiplication of networks is a very useful operation in network analysis, but it is dangerous when applied on large networks. The product of two large sparse networks need not to be sparse itself – it can "explode." In many interesting cases (e.g., genealogical and scientometric networks (Batagelj and Mrvar 2008; Batagelj and Cerinšek 2013)), the sparseness of the product can be guaranteed. A fast network multiplication procedure for sparse networks was added to **Pajek** in 2005.

In summer of 2005, while VB was visiting NICTA in Sydney, Australia, he was trying to analyze the two-mode IMDB network that was one of the tasks for that year's graph drawing contest. The two-mode hubs and authorities procedure did not give interesting results. There were no other appropriate tools in **Pajek**. So, VB asked AM to include in **Pajek** the counting of four rings and two-mode cores. The submitted solution (Ahmed et al. 2007) that was prepared in cooperation with NICTA's group won the first prize.

In 2007 a fast hierarchical method for clustering with relational constraints (Ferligoj and Batagelj 1983) was developed based on the idea to consider only the values on the existing lines.

Around 2005 the 64-bit PC computers started to appear. Since they can have memories over 4G, they are very important for further development of **Pajek**. The problem was that **Pajek** had to wait for 64-bit Delphi pascal compiler, and for several years, the answer was "mañana." Finally, in September 2011, the compiler became available.

Program **Pajek** was significantly reconstructed in the period 2011–2012: version **Pajek** 3 (32 and



Pajek and PajekXXL, Fig. 1 Snapshot of VR scene made with **Pajek** for graph drawing contest (1997)

64 bit) was shipped. **Pajek** 3 has a renewed and optimized menu structure. New community detection methods and layout algorithms for large networks were included. Additionally to "standard" **Pajek**, also version of **PajekXXL** with a compact data structure was finished in 2012. It is supposed to be used for huge networks (in which labeling of vertices is not important) that cannot be loaded into standard **Pajek**. **PajekXXL** can analyze networks with up to two billion vertices on everyday computers. Using **PajekXXL** we can identify and extract some smaller, interesting parts of a huge network that can be further analyzed (and visualized) with more sophisticated methods available in standard **Pajek**.

There exist several algorithms for finding communities but not all of them are suitable for analysis of very large networks. We decided to include in **Pajek** two community detection methods: Louvain method and VOS clustering. Louvain method is one of the most often used (and also among the fastest) methods. In its basic form, it was defined by Blondel et al. (2008) as a multilevel coarsening community detection method. In **Pajek** some improvements and extensions of this method suggested by Rotta and Noack (2011) were implemented. The second community detection method included in **Pajek** is VOS clustering (Waltman et al. 2010) that can be easily modified to be used also for visualization of networks – VOS mapping technique is included in **Pajek** as well. It is especially well suited for visualization of denser networks.

In the period 2014–2016, **Pajek** and **PajekXXL** were further optimized to be better compatible with the latest compilers and libraries. **Pajek** and **PajekXXL** version 4 was introduced. Some new options for analyzing acyclic networks were implemented, e.g., the main path analysis was generalized as suggested by Liu and Lu (2012) (key-Route searches). Several improvements were done on the visualization part. Some of them include:

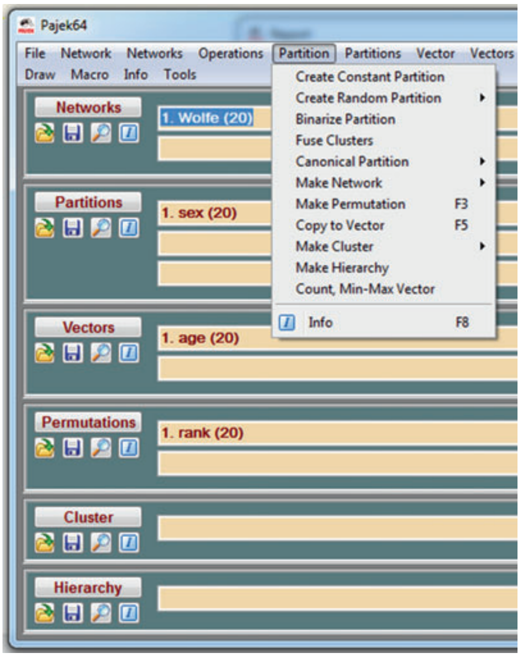
- Possibility to put the visualization window into a *Fisheye* mode was added (hovering the mouse pointer over the visualization window magnifies area around the mouse pointer).
- Showing labels of vertices on demand only – as tooltips/hints – was included (label appears only when mouse pointer touches the vertex).
- Assigning Unicode symbols to partition clusters and marking vertices with these symbols was added (see Pajek: using symbols in additional reading).
- Three additional shapes were added in EPS/SVG export ("house," "man," and "woman"). Some visualizations with additional vertex shapes are available in Mrvar and Batagelj (2016). See also Pajek: shapes of vertices in additional reading.
- Drawing vertices and lines transparently in SVG was enabled in **Pajek** version 4 as well.

In 2016 first steps in development of **Pajek3XL** were done. It is planned that **Pajek3XL** will be able to analyze networks where the number of vertices is larger than 2^{32} (in the first version, the limit is set to 10 billion).

For additional details on **Pajek**'s background, see **Pajek** history (2016) wiki page, and on **Pajek**'s evolution, see the history file at **Pajek**'s site.

Pajek

In **Pajek** analysis and visualization are performed using six data types:



- *Network* (graph)
- *Partition* (nominal or ordinal properties of vertices)
- *Vector* (numerical properties of vertices)
- *Cluster* (subset of vertices)
- *Permutation* (reordering of vertices, ordinal properties)
- *Hierarchy* (general tree structure on vertices)

The power of **Pajek** is based on several transformations that support different transitions among these data structures. Also the menu structure of the main **Pajek**'s window is based on them. Pajek's main window uses a "calculator" paradigm with a list accumulator for each data type. The operations are performed on the currently active (selected) data and are also returning the results through accumulators. The procedures are available through the main window menus. The menu options are organized with respect to the type(s) of input data. Scalars are treated as vectors of length 1.

Formats

Pajek has its own input data format (see the left side of Fig. 2). To describe a network, we first list the set of vertices followed by a list of directed lines or *arcs* and/or list of undirected lines or *edges*. The list of vertices starts with a keyword **`*vertices`** n where n is the total number of vertices in a network. Each vertex is described in its own line by its number (in the range $1 \ldots n$) followed by its label. Vertex descriptions can be omitted. The list of arcs starts with a keyword **`*arcs`**. Each arc is described with a triple u v w saying that there is an arc with the initial vertex u, the terminal vertex v, and the weight w. The weight w can be omitted – w gets the default value $w = 1$. Similarly in the list of edges, the triple u v w tells that there is an edge linking vertices u and v with the weight w. In the same network, we can mix arcs and edges and there can be parallel lines linking the same pair of vertices. **Pajek**'s format supports (BOM marked) Unicode. An input line starting with % is a comment.

Pajek supports also a compact description of line lists in the form of lists of vertex neighbors (**`*arcslist`** and **`*edgeslist`**).

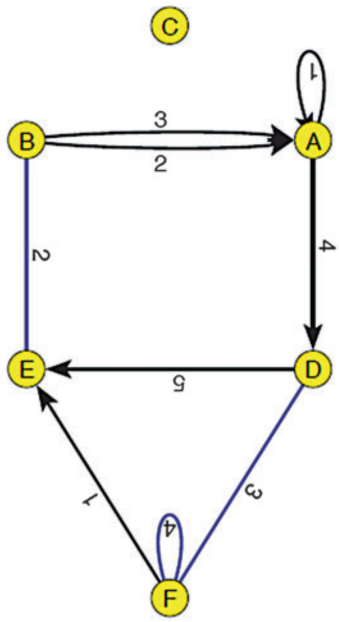
To enable easy input of older network data in a form of a matrix, the list(s) of lines in a network description can be replaced by a matrix – see the network description on the right side of Fig. 2.

Besides standard networks, **Pajek** supports also some extended types of networks: *two-mode networks*, bipartite (valued) graphs, networks between two disjoint sets of vertices; *multi-relational networks*, different relations on the same set of vertices; and *temporal networks*, dynamic graphs, networks changing over time.

For example, in the description of a two-mode network (see Fig. 3), the list of vertices starts with **`*vertices`** n k where k is the number of vertices in the first set. In the list of vertices, we first list all vertices from the first set, followed by all vertices in the second set. Each line has to have one end vertex in the first set and the other vertex in the second set.

Types of networks can be combined – for example, we can have a two-mode, multi-relational temporal network.

```
% example network
*vertices 6
1 "A"
2 "B"
3 "C"
4 "D"
5 "E"
6 "F"
*arcs
1 1 1
1 4 4
2 1 2
2 1 3
4 5 5
6 5 1
*edges
2 5 2
4 6 3
6 6 4
```

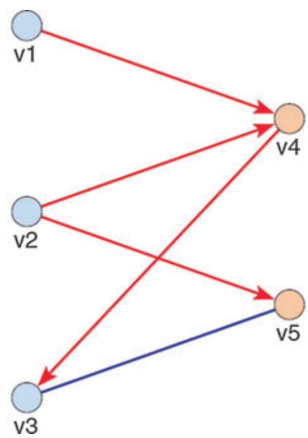



```
*vertices 6
1 "A"
2 "B"
3 "C"
4 "D"
5 "E"
6 "F"
*matrix
1 0 0 4 0 0
5 0 0 0 2 0
0 0 0 0 0 0
0 0 0 0 5 3
0 2 0 0 0 0
0 0 0 0 3 4
```

Pajek and PajekXXL, Fig. 2 Pajek input formats and picture of network

Pajek and PajekXXL, Fig. 3 Two-mode network and its description

```
*vertices 5 3
*arcs
1 4
2 4
2 5
4 3
*edges
3 5
```



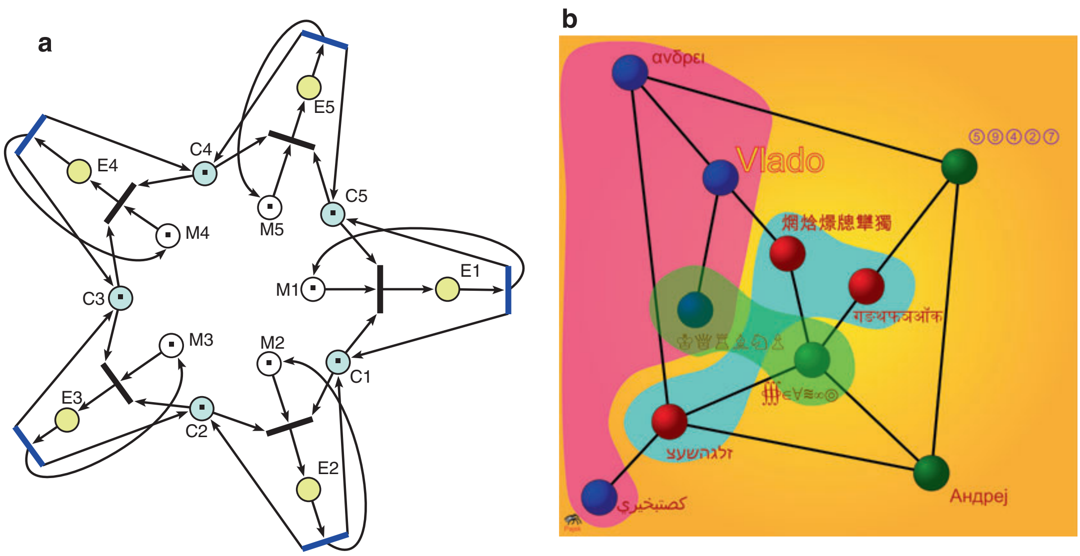
Pajek's input format allows also detailed specification of elements' (vertices and lines) graphical attributes. This enables the user to produce network layouts as is the Petri net presented on the left side of Fig. 4. For details see **Pajek**'s manual.

To describe a network, we have to provide (beside its graph structure) also data about properties of its vertices. All three types of property files have the same structure:

vertices** ***n
p_1
...
p_n

The ith vertex has value p_i with the following meaning:

- Vectors (VEC) – *numeric* data about vertices, $p_i \in \mathbb{R}$: the property has value p_i on vertex i.



Pajek and PajekXXL, Fig. 4 Petri net picture using **Pajek**'s file graphic attributes and an enhanced picture with Unicode labels

- Partitions or clusterings (CLU) – *nominal* or *ordinal* data about vertices, $p_i \in \mathbb{N}$: vertex i belongs to the cluster p_i.
- Permutations (PER) – *ordering* of vertices, $p_i \in \mathbb{N}$: vertex i is at the p_i-th position.

All types of data files can be combined into a single file – **Pajek**'s ***project*** file (PAJ).

Pajek can read data also in some other formats: UCINET DL files, genealogies in GEDCOM format, and molecular data in formats BS, MAC, and MOL.

Methods

This is a list of the main network analysis methods implemented in **Pajek**:

- Basic operations on **Pajek**'s structures (extraction, shrinking, combinations, conversions, arithmetics, etc.)
- Transforming temporal and multirelational networks into *collections* of networks (time slices, single-relation networks)
- Connectivities: weak, strong, biconnectivity, and periodic, condensation (Tarjan 1983; Harary et al. 1965)
- Shortest paths, k-neighbors, and flow
- Measures of vertex's importance: degrees, closeness, betweenness, and (corrected) clustering coefficient
- Kleinberg's hubs and authorities (also for two-mode networks) (Kleinberg 1998)
- McCabe software metrics (McCabe 2003)
- Structural holes (Burt 1992)
- Brokerage
- Vertex and line cuts
- Vertex and line islands (Zaveršnik and Batagelj 2004)
- (Generalized) cores (Batagelj and Zaveršnik 2011), two-mode cores (Ahmed et al. 2007), triadic spectrum; three-ring and four-ring weights (Batagelj and Mrvar 2001; Ahmed et al. 2007)
- Pathfinder skeleton (Schvaneveldt et al. 1988.
- Fragment (motif) searching
- Clustering of small networks
- (Generalized) blockmodeling of small networks (Doreian et al. 2000, 2004)
- Hierarchical clustering with relational constraints (Ferligoj and Batagelj 1983)
- Community detection methods (Blondel et al. 2008; Waltman et al. 2010)
- Methods for partitioning signed networks (Doreian and Mrvar 1996)

- Basic operations on acyclic networks (depth, topological ordering, CPM)
- Kinship analysis (White et al. 1999; Batagelj and Mrvar 2008)
- Hummon-Doreian weights in acyclic networks (Hummon and Doreian 1989; Batagelj 2003)
- Computing probabilistic flow in acyclic networks
- Generalized main path analysis of acyclic networks (so called key-Route searches as defined by Liu and Lu 2012)
- Multiplication of networks (Batagelj and Mrvar 2008; Batagelj and Cerinšek 2013)
- Normalizations of weighted networks (Batagelj and Mrvar 2003)
- Basic support for Petri nets (Peterson 1981)
- Generation of different types of random networks (small world, scale-free, etc.) (Batagelj and Brandes 2005; Pennock et al. 2002)
- Computing different network indices (e.g., modularity, assortativity, relinking index, E-I index, etc.)

The "granularity" of **Pajek**'s methods is high – usually we need to perform a sequence of operations to achieve the intended result.

Visualization

Pajek's layout is basically in a 3D unit cube $x, y, z \in [0, 1]$. In 2D layouts, the third dimension is not considered.

Pajek provides a collection of general graph drawing algorithms such as Kamada-Kawai, Fruchterman-Reingold, VOS, and MDS.

In **Pajek** vectors can be displayed as *size* (width and height) of vertex (figure), as its *coordinate*, and partitions as *color* or *shape* of the figure, *Unicode symbol*, or as a vertex *label* (content, font size, and color).

The weights on lines can be displayed as *value*, *thickness*, or *gray level*. Nominal line values can be assigned as *label*, *color*, or *line pattern* (see Pajek manual, Sect. 4.3).

The layouts can be exported in different formats: bitmap, JPEG, EPS, SVG, X3D, MOL, and Kinemages. Pictures of network in EPS or SVG can be imported in vector graphics programs such as Inkscape, Adobe Illustrator, or CorelDRAW. They can be used to enhance the pictures (see the right side of Fig. 4). Nice examples of this approach are visualizations produced by FAS Research, Vienna. Darko Brvar developed a program **SVGanim** for dynamic visualization of temporal networks.

The important parts of network identified by **Pajek**'s methods are often dense and the traditional dots and lines layout is unreadable. In such cases, a better solution is the *matrix display* for the "right" ordering of vertices (usually determined with clustering or blockmodeling). In Fig. 5 the network of collaboration between countries in EU projects on simulation is presented.

A natural display of an acyclic network is by layers – all arcs are pointing in the same direction (up or down). In Fig. 6 the main island in the citation network on "social networks" is presented – it is an acyclic network.

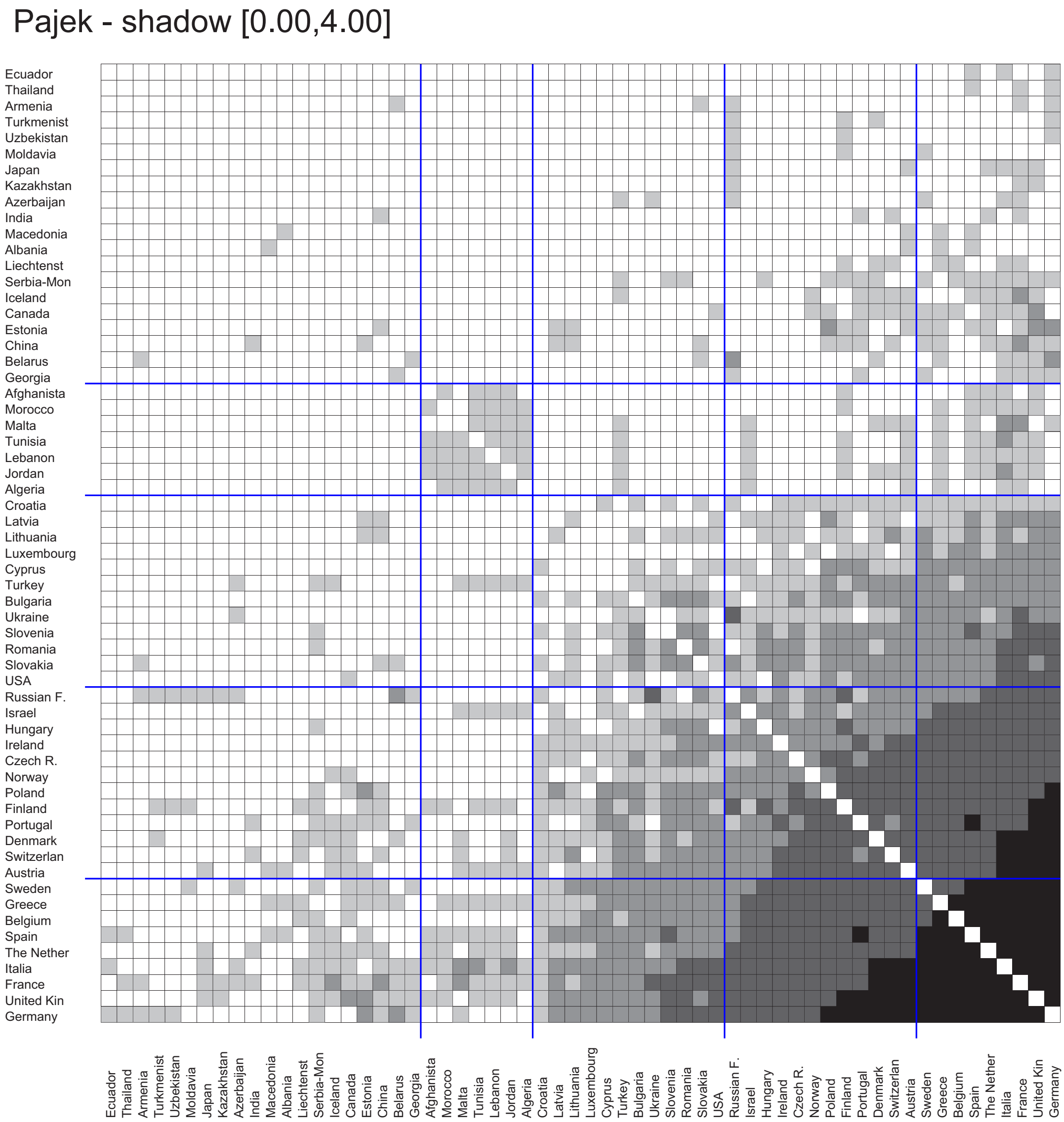
Interoperability

In the development of **Pajek**, we often encountered the need to include in **Pajek** different statistical procedures. We decided not to do this, but to provide a connection to statistical programs such as R, SPSS, or Excel. Other statistical programs can import results obtained by **Pajek** via export to tab-delimited files. The menu option **Tools** allows the user to connect **Pajek** with other programs – for example, viewers. This is also the way in which **PajekXXL** calls **Pajek** after the large network is decomposed to smaller (manageable) parts which can be analyzed and visualized with more sophisticated methods which are available in **Pajek** (but not in **PajekXXL**).

Many network analysis programs support (at least partially) **Pajek**'s input format. There exist also programs to transform specific types of data into **Pajek**'s network data – for example, Text2Pajek and WoS2Pajek.

Special Options

In **Pajek** frequently used sequences of operations can be defined and saved as *macros*. This allows also the adaptations of **Pajek** to groups of users from different areas (social networks, chemistry, genealogy, computer science, mathematics, etc.)



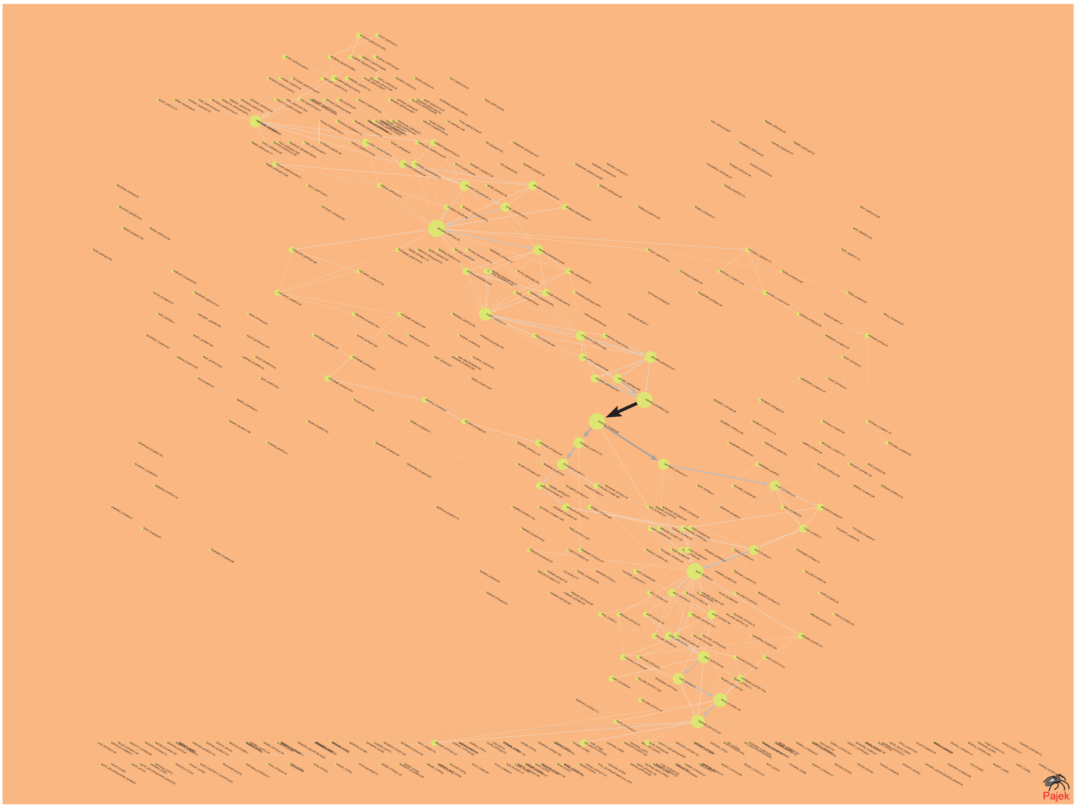
Pajek and PajekXXL, Fig. 5 Collaboration between countries in EU projects on simulation

for specific tasks. **Pajek** supports also *repetitive operations* on series (collections) of networks and/or other data objects.

Documentation

All the capabilities of **Pajek** are concisely described in its *manual* that is available at **Pajek**'s download page. A friendly introduction into SNA and corresponding **Pajek**'s commands is provided in the *book ESNA – Exploratory Social Network Analysis with Pajek* (De Nooy et al. 2011) – that was published also in Japanese and Chinese.

Basic info about **Pajek** and introduction to **Pajek** are available in different languages on the main **Pajek**'s page (e.g., Greek, Polish, Spanish, Portuguese, German, Persian, French, Japanese, Chinese, Italian, English, Slovenian). Several video lectures on using **Pajek** can be found as well.



Pajek and PajekXXL, Fig. 6 Main island in SN5

Users can also join **Pajek**'s *mailing list* and discuss their problems with other users (http://list.fmf.uni-lj.si/cgi-bin/mailman/listinfo/pajek).

Key Applications

Pajek was used in many different fields. In September 2017, the book *ESNA* had in Google Scholar over 3000 citations, and other papers about Pajek had over 3200 citations. As examples we present some applications in analysis of genealogies, in bibliometrics, and in analysis of hyperlink graphs.

Genealogies

Pajek was successfully applied to analysis of large genealogies. When reading GEDCOM files, Pajek can produce three types of networks: Ore graphs, p-graphs, and bipartite p-graphs. Each representation has some advantages for special uses.

Many interesting subnetworks can be found in genealogies, e.g., searching for the shortest genealogical paths between people, longest patrilineages, different statistics on marriages, and births. Batagelj and Mrvar (2008) identified 16 fragments that represent all possible relinking marriages (blood marriages or not) on at most six vertices in p-graphs. Using general fragment searching in **Pajek**, we obtain frequency distributions of these fragment counts which can be used to compare different genealogies. Relinking index as a measure of relinking by marriages among persons belonging to the same families was defined in the paper as well.

Pajek generates three relations when reading genealogies as Ore graphs: . . .is a mother of. . ., . . .is a father of. . ., and . . .is a spouse of. . .. Using these three relations and network multiplication in **Pajek**, we can obtain other kinship relations (…is an aunt of…, …is a grandfather of…, …is a niece of…, etc.) as derived networks.

VB and AM collaborate with other researchers on analysis of genealogies: Klaus Hamberger (EHESS) is developing a specialized program PUCK for searching for matrimonial circuits in kinship networks (Hamberger 2016). The program is compatible with **Pajek**. **Pajek** is intensively used also by Douglas White (UCI). On his web site, we can find several references for using **Pajek** in anthropological research (White 2016).

Bibliographic Data

Pajek is a powerful tool for analyzing all kinds of bibliographical data (Batagelj and Mrvar 2000; Batagelj and Cerinšek 2013). Loet Leydesdorff produced several specialized programs for creation and analysis of bibliographical data that can be well combined with **Pajek** (Leydesdorff 2016).

Pajek is also well connected to program VOSviewer (VOSviewer 2017). VOSviewer is primarily intended to be used for analyzing bibliometric networks. As explained in previous sections, there are several algorithms in **Pajek** that produce partitions and vectors. **Pajek** can export networks, partitions, and vectors directly to VOSviewer and use its additional visualization techniques.

Hyperlink Graph

PajekXXL was successfully applied to analysis of huge hyperlink networks. It was used in Web Data Commons – Hyperlink Graphs project (network containing 43 million vertices and 623 million arcs). For details see http://webdatacommons.org/hyperlinkgraph/index.html.

Future Directions

Besides new algorithmic development, there are many things still missing in **Pajek** and waiting for implementation. For example:

- Further increasing the upper limit of vertices that **PajekXXL** can handle, developing **Pajek3XL**
- GUI control of elements' graphical attributes
- Internationalization of GUI and messages
- Simple programming language for macros (with variables and control structure)
- Support for vertex idents and namespaces
- More interactive visualization
- Support for additional input/output formats (GraphML, JSON, HTML 5, D3.js, etc.)
- Computations with weights based on semirings (Batagelj 1994)
- Genealogies of data
- Layout styles
- Support for analysis of temporal networks (Batagelj and Praprotnik 2016)

Cross-References

▶ Network Data File Formats
▶ Semirings and Matrix Analysis of Networks
▶ Sources of Network Data
▶ UCINET
▶ Visualization of Large Networks

Acknowledgments The work was supported in part by the ARRS, Slovenia, grant P1-0294, as well as by grant N1-0011 within the EUROCORES Programme EUROGIGA (project GReGAS) of the European Science Foundation.

References

Ahmed A, Batagelj V, Fu X, Hong S-H, Merrick D, Mrvar A (2007) Visualisation and analysis of the Internet movie database. In: Asia-Pacific symposium on visualisation 2007, Sydney, Australia, 5–7 February 2007 (IEEE Cat. No. 07EX1615), pp 17–24

Batagelj V (1987) Data structure graph. Eight Yugoslav seminar on graph theory, Novi Sad, 17–18 Apr 1987. Institute of Mathematics, Novi Sad, p 4

Batagelj V (1994) Semirings for social networks analysis. J Math Sociol 19(1):53–68

Batagelj V (1997) Notes on blockmodeling. Soc Netw 19:143–155

Batagelj V (2003) Efficient algorithms for citation network analysis. http://arxiv.org/abs/cs.DL/0309023

Batagelj V, Brandes U (2005) Efficient generation of large random networks. Phys Rev E 71(3 Part 2):036113, 1–5
Batagelj V, Cerinšek M (2013) On bibliographic networks. Scientometrics 96(3):845–864
Batagelj V, Mrvar A (1998) Pajek: a program for large network analysis. Connections 21(2):47–57
Batagelj V, Mrvar A (2000) Some analyses of Erdös collaboration graph. Soc Netw 22:173–186
Batagelj V, Mrvar A (2001) A subquadratic triad census algorithm for large sparse networks with small maximum degree. Soc Netw 23(3):237–243
Batagelj V, Mrvar A (2003) Density based approaches to network analysis: analysis of Reuters terror news network. In: Workshop on link analysis for detecting complex behavior (LinkKDD2003), Washington, DC, 27 Aug 2003. http://www-2.cs.cmu.edu/~dunja/LinkKDD2003/papers/Batagelj.pdf
Batagelj V, Mrvar A (2008) Analysis of kinship relations with Pajek. Soc Sci Comput Rev 26(2):224–246
Batagelj V, Praprotnik S (2016) An algebraic approach to temporal network analysis based on temporal quantities. Soc Netw Anal Min 6(1):1–22
Batagelj V, Zaveršnik M (2007) Short cycle connectivity. Discret Math 307(3–5):310–318
Batagelj V, Zaveršnik M (2011) Fast algorithms for determining (generalized) core groups in social networks. Adv Data Anal Classif 5(2):129–145
Batagelj V, Doreian P, Ferligoj A (1992a) An optimizational approach to regular equivalence. Soc Netw 14(1–2):121–135
Batagelj V, Ferligoj A, Doreian P (1992b) Direct and indirect methods for structural equivalence. Soc Netw 14(1–2):63–90
Batagelj V, Mrvar A, Zaveršnik M (1999) Partitioning approach to visualization of large graphs. In: Graph drawing, 1999, Stiřin Castle, Czech Republic, September 1999. Lecture notes in computer science, vol 1731/1999. Springer, Berlin, Heidelberg, pp 90–97
Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp. https://doi.org/10.1088/1742–5468/2008/10/P10008
Brusco M, Doreian P, Mrvar A, Steinley D (2011) Two algorithms for relaxed structural balance partitioning: linking theory, models, and data to understand social network phenomena. Sociol Methods Res 40:57–87
Burt RS (1992) Structural holes. The social structure of competition. Harvard University Press, Cambridge
Cormen T, Leiserson C, Rivest R (1990) Introduction to algorithms. McGraw-Hill, New York
Doreian P, Mrvar A (1996) A partitioning approach to structural balance. Soc Netw 18:149–168
Doreian P, Mrvar A (2009) Partitioning signed social networks. Soc Netw 31(1):1–11
Doreian P, Mrvar A (2014) Testing two theories for generating signed networks using real data. Adv Methodol Stat 11:31–63. Faculty of Social Sciences, Ljubljana
Doreian P, Mrvar A (2015) Structural balance and signed international relations. J Soc Struct 16. Carnegie Mellon University, Pittsburgh, 1–49
Doreian P, Mrvar A (2016) Identifying fragments in networks for structural balance and tracking the levels of balance over time. Connections 35(2):6–18
Doreian P, Batagelj V, Ferligoj A (2000) Symmetric-acyclic decompositions of networks. J Classif 17(1):3–28
Doreian P, Batagelj V, Ferligoj A (2004) Generalized blockmodeling of two-mode network data. Soc Netw 26(1):29–53
Doreian P, Lloyd P, Mrvar A (2013) Partitioning large signed two-mode networks: problems and prospects. Soc Netw 35:178–203
Ferligoj A, Batagelj V (1983) Some types of clustering with relational constraints. Psychometrika 48(4):541–552
Hamberger K (2016) Puck – program for the use and computation of kinship data. http://www.kintip.net
Harary F, Norman RZ, Cartwright D (1965) Structural models: an introduction to the theory of directed graphs. Wiley, New York
Hill RA, Dunbar RIM (2002) Social network size in humans. Hum Nat 14(1):53–72
Hummon NP, Doreian P (1989) Connectivity in a citation network: the development of DNA theory. Soc Netw 11:39–63
Kleinberg J (1998) Authoritative sources in a hyperlinked environment. In: Proceedings of 9th ACM-SIAM symposium on discrete algorithms, San Francisco, pp 668–677. http://www.cs.cornell.edu/home/kleinber/auth.ps
Leydesdorff L (2016) Web site: http://www.leydesdorff.net/
Liu JS, Lu LYY (2012) An integrated approach for main path analysis: development of the Hirsch index as an example. J Am Soc Inf Sci Technol 63:528–542
McCabe T (2003) Computer science approaches: visualization tools and software metrics. In: Survey automation. NAP, Washington, DC, pp 116–136
Mrvar A, Doreian P (2009) Partitioning signed two-mode networks. J Math Sociol 33(3):196–221
Pennock DM et al (2002) Winners don't take all. PNAS 99(8):5207–5211
Peterson JL (1981) Petri net theory and the modeling of systems. Prentice-Hall, Englewood Cliffs
Rotta R, Noack A (2011) Multilevel local search algorithms for modularity clustering. J Exp Algorithmics 16:2.3
Schvaneveldt RW, Dearholt DW, Durso FT (1988) Graph theoretic foundations of pathfinder networks. Comput Math Appl 15(4):337–345
Seidman SB (1983) Network structure and minimum degree. Soc Netw 5:269–287
Tarjan RE (1983) Data structures and network algorithms. Society for Industrial and Applied Mathematics, Philadelphia
VOSviewer web site: http://www.vosviewer.com/

Waltman L, van Eck NJ, Noyons ECM (2010) A unified approach to mapping and clustering of bibliometric networks. J Informetr 4:629–635

White D (2016) Web site: http://eclectic.ss.uci.edu/~drwhite/

White DR, Batagelj V, Mrvar A (1999) Analyzing large kinship and marriage networks with graph and Pajek. Soc Sci Comput Rev 17(3):245–274

Zaveršnik M, Batagelj V (2004) Islands, slides from Sunbelt XXIV, Portorož, 12–16 May 2004. http://vlado.fmf.uni-lj.si/pub/networks/doc/sunbelt/islands.pdf

Recommended Reading

Batagelj V (2009a) Complex networks, visualization of. In: Meyers RA (ed) Encyclopedia of complexity and systems science. Springer, New York/London, pp 1253–1268

Batagelj V (2009b) Social network analysis, large-scale. In: Meyers RA (ed) Encyclopedia of complexity and systems science. Springer, New York/London, pp 8245–8265

Batagelj V (2011) Large-scale network analysis. In: Scott J, Carrington P (eds) The SAGE handbook of social network analysis. SAGE, Thousand Oaks

Batagelj V, Mrvar A (2003) Pajek – analysis and visualization of large networks. In: Juenger M, Mutzel P (eds) Graph drawing software. Mathematics and visualization. Springer, Berlin, pp 77–103

Batagelj V, Doreian P, Ferligoj A, Kejžar N (2014) Understanding large temporal networks and spatial networks: exploration, pattern searching, visualization and network evolution. Wiley, Hoboken, New Jersey

De Nooy W, Mrvar A, Batagelj V (2011) Exploratory social network analysis with Pajek, revised and expanded 2nd edn. Structural analysis in the social sciences. Cambridge University Press, New York, Sept 2011. Translation in Japanese 2008, Denkyo Press, Tokyo. Translation in Chinese 2012, Beijing World Publishing Corporation

Doreian P, Batagelj V, Ferligoj A (2004) Generalized blockmodeling. Structural analysis in the social sciences. Cambridge University Press, Cambridge

Mrvar A, Batagelj V (2016) Analysis and visualization of large networks with program package Pajek. Complex Adapt Syst Model 4:6. SpringerOpen

Pajek web site. http://mrvar.fdv.uni-lj.si/pajek

Pajek wiki. http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek

Pajek: historical background. http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:histback

Pajek: shapes of vertices. http://mrvar.fdv.uni-lj.si/pajek/Transparent/ManWomanHouse.htm

Pajek: using symbols. http://mrvar.fdv.uni-lj.si/pajek/Symbols/symbols-examples.htm

Pajek's datasets. http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:data:index

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge