Complex Adaptive Systems Modeling

Analysis and visualization of large networks with program package Pajek

Andrej Mrvar[1*] and Vladimir Batagelj[2]

*Correspondence:
Andrej.Mrvar@fdv.uni-lj.si
[1] Faculty of Social
Sciences, University
of Ljubljana, Kardeljeva pl. 5,
1000 Ljubljana, Slovenia
Full list of author information
is available at the end of the
article

Abstract

Pajek is a program package for analysis and visualization of large networks (networks containing up to one billion of vertices, there is no limit—except the memory size—on the number of lines). It has been available for 20 years. The program, documentation and supporting material can be downloaded and used for free for noncommercial use from its web page: http://mrvar.fdv.uni-lj.si/pajek/.

Keywords: Pajek, PajekXXL, Analysis of large networks, Visualization of networks

Background

Large networks can be found everywhere, e.g., social networks, connections among people (kinship relations, friendship, *Facebook, Twitter, WWW*); trade among organizations or countries; citation and co-authorship networks (e.g., obtained from *Web of Science*); telephone calls; flow charts in computer science; organic molecule in chemistry (e.g., DNA, protein–protein interaction networks, genome research); connections among words in text or dictionaries; transportation networks (airlines, streets,…).

Review

History

Development of program Pajek began in 1996 when Andrej Mrvar started his work on Ph.D. thesis on analysis and visualization of large networks at Faculty of Computer and Information Science, University Ljubljana (advisor prof. dr. Vladimir Batagelj). Although Pajek has been developed now for 20 years, it is still the only general program available on the market that can handle huge networks (networks having up to a billion of vertices; there is no limit—except the memory size—on the number of lines). Pajek is now used by several universities (e.g., *University of Oxford, University of California at Irvine, San Diego, Amsterdam…*) and companies (e.g., *Deutsche Bundesbank, Volkswagen AG, SPSS Korea, Bank of England, Cisco, Basel Bank for International Settlements, Kansas City Missouri Police Department, Indianapolis Police Department…*). Pajek is cited (March, 2016) more than 2000 times in *Web of Science* and more than 6000 times in *Google Scholar*. Citations can be found also in some prominent journals like *Nature* and *Genome Research*. Pajek was the winner of *the William D. Richards Jr., Software Award* in 2013.

Main goals of Pajek

In Pajek analysis and visualization of large networks are performed using six data types (objects): *network* (graph); *partition* (nominal or ordinal properties of vertices); *vector* (numerical properties of vertices); *cluster* (subset of vertices); *permutation* (reordering of vertices, ordinal properties); and *hierarchy* (general tree structure on vertices). In this way Pajek main window is organized (see Fig. 1 for the snapshot of the Pajek main window).

The main goals in the design of Pajek are:

- to support *abstraction* by (recursive) decomposition of a large network into several smaller networks that can be treated further using more sophisticated methods (Batagelj and Mrvar 1998; Batagelj et al. 1999);
- to provide the user with some powerful *visualization* tools (Batagelj and Mrvar 2002, 2003);
- to implement a selection of *efficient (subquadratic) algorithms* for analysis of large networks (Batagelj and Mrvar 2014).

Operations

According to main goals, Pajek contains several basic operations on its objects. Pajek is not '*a one click program*', some users call it '*the network calculator*'. That means that for obtaining some result several basic operations must be executed in a sequence. In fact possibility to combine different basic operations gives Pajek a special power.

Some of such basic operations available in Pajek include: extracting subnetworks; shrinking selected parts of networks; searching for connected components (weak, strong, biconnected); searching for shortest paths, k-neighbors, maximum flow; computing centralities of vertices and centralizations of networks (degree, closeness, betweenness, hubs and authorities, clustering coefficients, Laplacean centrality); fragment searching;
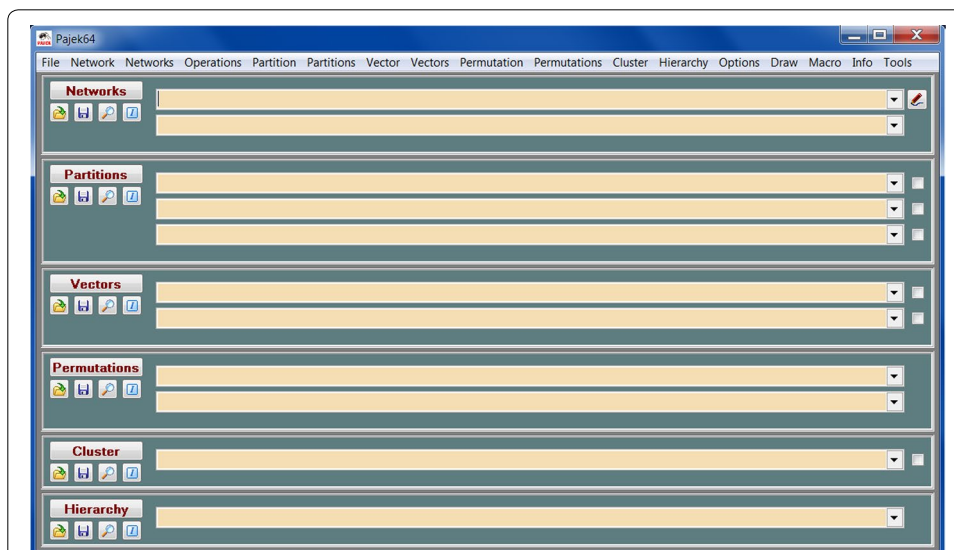
Fig. 1 Pajek main window

clustering in networks (with or without constraint); brokerage; community detection methods (Louvain method and VOS clustering); fast sparse network multiplication; triadic census (Batagelj and Mrvar 2001); structural holes; islands (on vertices or lines); counting for three and four rings; generating different types of random networks, Petri nets execution, and many, many others. Some operations which are suitable only for smaller networks are also included in Pajek (e.g., Pajek is the only software that includes generalized blockmodeling (Batagelj et al. 2004)—a sophisticated method for partitioning smaller networks).

Pajek contains several operations that enable transitions among objects. In this way analysis of large networks can be performed fluently. Sequences of commands can be defined as a macro.

Results obtained by Pajek (e.g., partitions and vectors) can be further analysed using *R, SPSS*, and *Excel* (several exports to statistical software are included in Pajek).

In addition to ordinary (directed or undirected) networks Pajek supports also *two-mode networks, temporal networks* (networks changing over time), *signed networks* (networks with positive and negative lines), *multirelational networks* (several relations defined on the same set of vertices) and *acyclic networks*. Special operations for these kinds of networks are available, e.g., different methods for partitioning signed networks (Doreian and Mrvar 1996, 2009, 2014, 2015; Mrvar and Doreian 2009; Doreian et al. 2013); several methods for computing traversal weights (e.g., *SPC, SPLC,* and *SPNP*) and later determining main paths in acyclic (e.g., citation) networks….

Genealogies saved in GEDCOM format can be loaded in Pajek as well. Kinship relations can be represented as *Ore graph, p-graph* or *bipartite p-graph* (White et al. 1999; Batagelj and Mrvar 2008). Pajek was successfully applied to analysis of large genealogies (e.g., searching for relinking marriages).

In addition to standard Pajek there exists also a special version called *PajekXXL.* PajekXXL is a special edition of program Pajek which memory consumption is much lower. For the same sparse network it needs at least 2–3 times less physical memory than Pajek. Operations that are memory intensive (e.g., generating random networks, extracting, shrinking…) are therefore much faster. PajekXXL is usually used for huge networks that do not fit to available computer memory. After some interesting parts are found and extracted, the standard Pajek version can be used for further analysis.

Visualization methods

There are different methods for automatic generation of network layouts available in Pajek. The most important (often used) are: Kamada-Kawai optimization, Fruchterman Reingold optimization, VOS mapping, Pivot MDS, drawing in layers, FishEye transformation. Layouts obtained by Pajek can be exported to different 2D or 3D output formats (e.g., SVG, EPS, X3D, VOSViewer, Mage,…). Special viewers and editors for these formats are available (e.g., inkscape, GSView, instantreality, KiNG,…). Using them we can further edit layouts or examine them in details.

Pajek implementations of the algorithms for automatic network drawing were tested on several *Graph Drawing Competitions.* Andrej Mrvar and Vladimir Batagelj took part in competitions in the period 1995–2005. Their visualizations using program Pajek were awarded altogether with eight first and three second prizes.

Publications

Because of such a wide usage of the program the need for a monograph describing 'How to do network analysis with Pajek?' became inevitable.

The monograph: *Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj: Exploratory Social Network Analysis with Pajek* was published by *Cambridge University Press* in 2005 (de Nooy et al. 2005). Around 6000 copies of the first edition of the monograph were sold. Around 2500 citations of this publication can be found in Google Scholar. Because of that the Cambridge University Press published also the second-revised and expanded-edition in 2011 (de Nooy et al. 2011). The monograph was translated to Japanese (Pajek を活用した社会ネットワーク分析) and published by *Tokyo Denki University Press* in 2009 (de Nooy et al. 2009). It was translated to Chinese by *Beijing World Publishing Corporation* in 2012 as well (蜘蛛：社会网络分析技术). In the meantime the first China edition was already sold and the second Chinese edition was published (de Nooy et al. 2012). See Fig. 2 for cover pages with links to all five book editions.

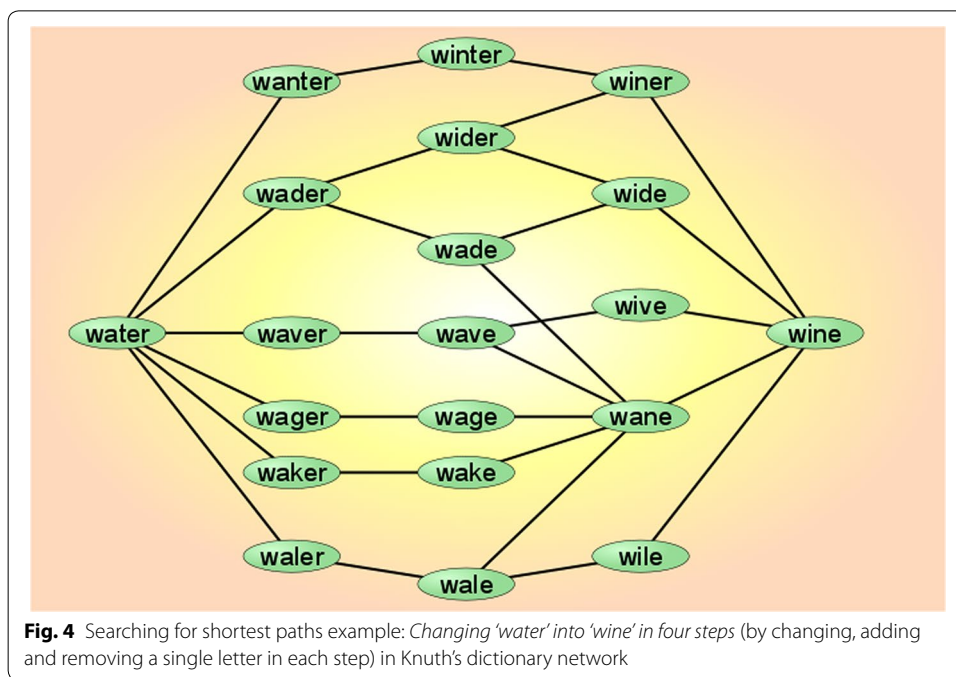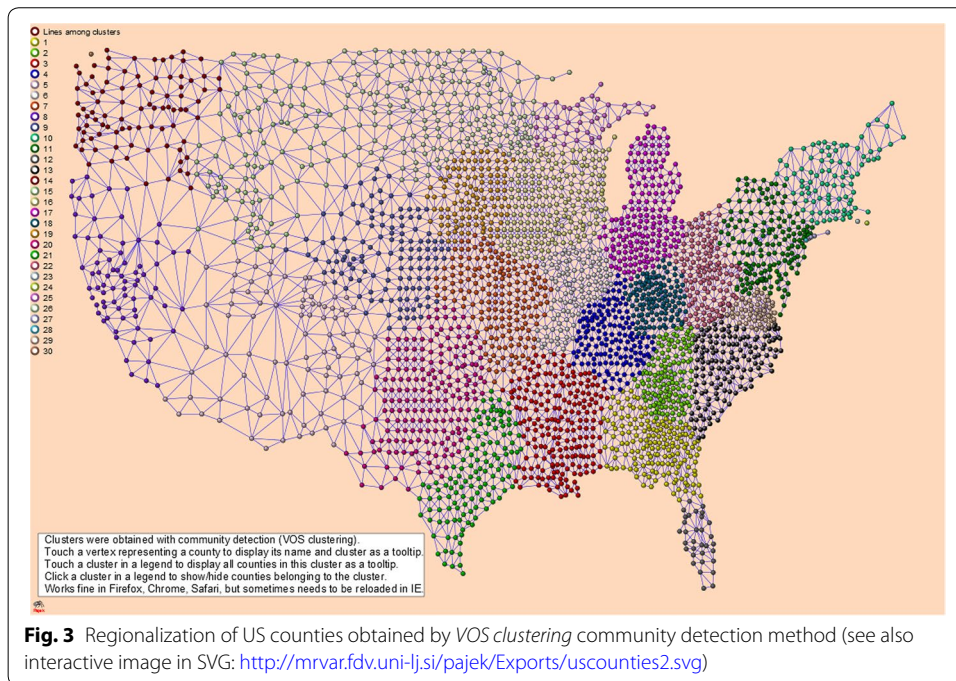## Example visualizations obtained by Pajek

In Figs. 3, 4, 5, 6 and 7 some typical examples of analysis and visualization of large networks using Pajek are shown. Several more examples can be found on the Pajek web page: http://mrvar.fdv.uni-lj.si/pajek/.
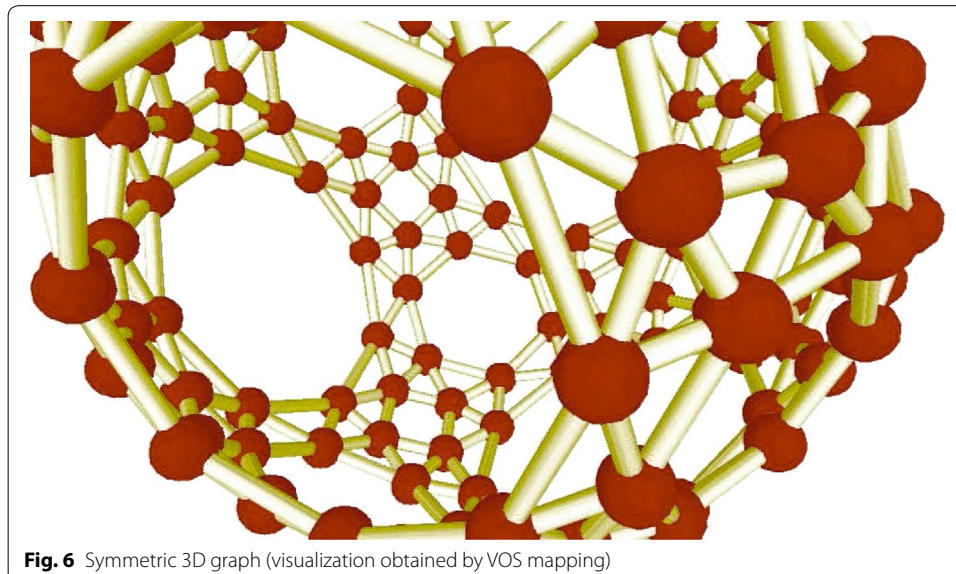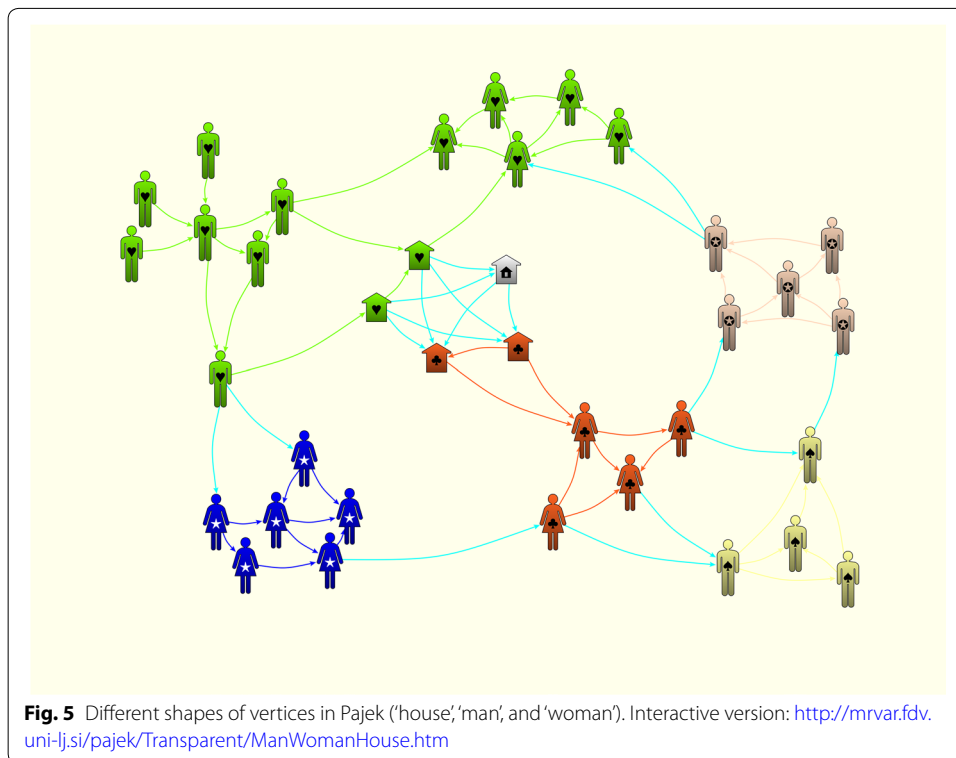

Fig. 2 Links to all five book editions. 1. First English Edition, 2. Second English Edition, 3. Japanese Edition, 4. First Chinese Edition, 5. Second Chinese Edition
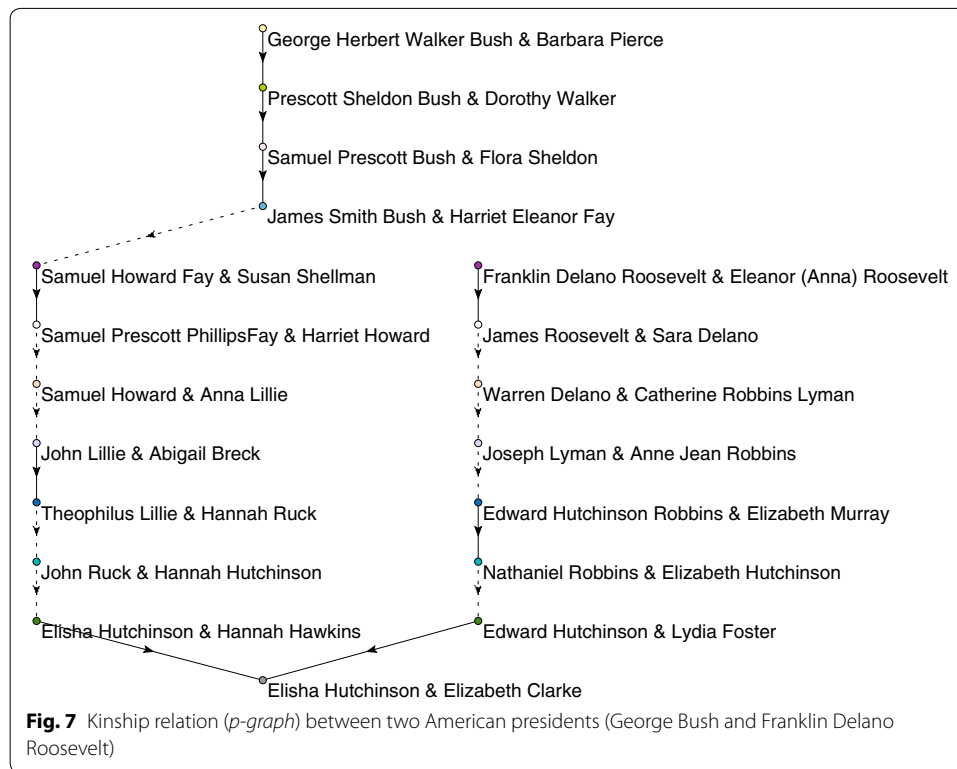
Discussion

One of the implications of the 20 years development in information and communication technologies is availability of huge amount of data. Some people call this phenomenon Big Data. Now a lot of data is already available in a computer readable electronic

Fig. 3 Regionalization of US counties obtained by *VOS clustering* community detection method (see also interactive image in SVG: http://mrvar.fdv.uni-lj.si/pajek/Exports/uscounties2.svg)

Fig. 4 Searching for shortest paths example: *Changing 'water' into 'wine' in four steps* (by changing, adding and removing a single letter in each step) in Knuth's dictionary network

Fig. 5 Different shapes of vertices in Pajek ('house', 'man', and 'woman'). Interactive version: http://mrvar.fdv. uni-lj.si/pajek/Transparent/ManWomanHouse.htm

Fig. 6 Symmetric 3D graph (visualization obtained by VOS mapping)

form. Researchers have recognized the power of gathering and analyzing such data. The need for computer programs to analyze and visualize Big Data becomes inevitable. The same happens in the area of social networks analysis. Now we can generate huge networks from different electronic resources. Typical examples of such networks are coauthorship

Fig. 7 Kinship relation (*p-graph*) between two American presidents (George Bush and Franklin Delano Roosevelt)

(Batagelj and Mrvar 2000) and citations networks obtained from Web of Science. Pajek—program for analysis and visualization of large networks is discussed in the paper.

Application areas

As mentioned in previous sections large networks could be found anywhere. Program package Pajek has several application areas, including: analysis of any kind of social networks (Facebook, Twitter, networks in organizations (Škerlavaj et al. 2010), networks of international relations, kinship relations,…); citation and co-authorship networks; protein−protein interaction networks; transportation networks, archaeological networks,… Some people use it also only as a visualization tool for displaying any kind of networks.

Conclusion

Pajek is in constant development. For more information on Pajek check its webpage. For those who want to learn Pajek: Several books, manuals, articles, and samples are available. They are listed in the references section.

Author details
[1] Faculty of Social Sciences, University of Ljubljana, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia. [2] Institute of Mathematics, Physics and Mechanics, University of Ljubljana, Jadranska 21, 1000 Ljubljana, Slovenia.

Competing interests
The authors declare that they have no competing interests.

References

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 Erdos 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 23:237–243

Batagelj V, Mrvar A (2002) Pajek—analysis and visualization of large networks. Lecture notes in computer science vol 2265, Springer-Verlag, pp 477–478

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

Batagelj V, Mrvar A (2008) Analysis of kinship relations with Pajek. Soc Sci Comput Rev 26(2):224–246

Batagelj V, Mrvar A (2014) Pajek. In: R Alhajj, J. Rokne (eds) Encyclopedia of social network analysis and mining. Springer, Heidelberg, pp 1245–1256

Batagelj V, Mrvar A, Zaveršnik M (1999) Partitioning Approach to visualization of large graphs. Lecture notes in computer science vol 1731. Springer-Verlag, pp 90–97

Batagelj V, Mrvar A, Ferligoj A, Doreian P (2004) Generalized blockmodeling with Pajek. Advances in methodology and statistics, vol 1. FDV, Ljubljana, pp 455–467

de Nooy W, Mrvar A, Batagelj V (2005) Exploratory social network analysis with Pajek. Cambridge University Press, New York

de Nooy W, Mrvar A, Batagelj V (2009) *Pajek* を活用した社会ネットワーク分析, Tokyo Denki University Press, Tokyo

de Nooy W, Mrvar A, Batagelj V (2011) Exploratory social network analysis with Pajek: revised and expanded, 2nd edn. Cambridge University Press, New York

de Nooy W, Mrvar A, Batagelj V (2012) 蜘蛛：社会网络分析技术. Beijing World Publishing Corporation, Beijing

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 31:1–11

Doreian P, Mrvar A (2014) Testing two theories for generating signed networks using real data. Advances in Methodology and Statistics, vol 11. FDV, Ljubljana, pp 31–63

Doreian P, Mrvar A (2015) Structural balance and signed international relations. J Soc Struct 16

Doreian P, Lloyd P, Mrvar A (2013) Partitioning large signed two-mode networks: problems and prospects. Soc Netw 35:178–203

Mrvar A, Doreian P (2009) Partitioning signed two-mode networks. J Math Sociol 33:196–221

Škerlavaj M, Dimovski V, Mrvar A, Pahor M (2010) Intra-organizational learning networks within knowledge-intensive learning environments. Interact Learn Environ 18:39–63

White DR, Batagelj V, Mrvar A (1999) Analysis of kinship relations with Pajek. Soc Sci Comput Rev 17(3):245–274