Appendices of: ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization

Mathieu Jacomy, *Non member* Sebastien Heymann, *Non member*
Tommaso Venturini, *Non member* and Mathieu Bastian, *Non member*

MM. Jacomy and Venturini are with Sciences Po, medialab.

M. Heymann is with LIP6 - CNRS - Universite Pierre et Marie Curie.

M. Bastian is with the Gephi Consortium.

Abstract

ForceAtlas2 is a force vector algorithm proposed in the Gephi software, appreciated for its simplicity and for the readability of the networks it helps to visualize. This paper presents its distinctive features, its energy-model and the way it optimizes the "speed versus precision" approximation to allow quick convergence. We also claim that ForceAtlas2 is handy because the force vector principle is unaffected by optimizations, offering a smooth and accurate experience to users.

Index Terms

Force Vector, Gephi, Spatialization, Network Visualization, Social Networks Analysis

I. Appendices

A. Datasets

The * datasets are available to download in the Gephi Datasets page[1].

- SiSWordsFR: 6704 french words linked by 71744 co-occurrence relations. Extracted from the Exalead search engine by Mathieu Jacomy, Franck Ghitalla, Sophie Tocreau and Martine Roussel for the "Ministere de l'Enseignement Superieur et de la Recherche" in 2007. This network is strongly clustered. This dataset is used in figures 1, 4, 5 and 13.
- LesMiserables*: A directed and weighted graph of 77 nodes and 254 edges. It represents the characters in the novel Les Miserables. This dataset is included in Gephi. Reference: D. E. Knuth, The Stanford GraphBase: A Platform for Combinatorial Computing, Addison-Wesley, Reading, MA (1993).
- KarateClub*: 34 nodes and 78 edges. Social network of friendships between members of a karate club at a US university in the 1970s. Reference: W. W. Zachary, An information flow model for conflict and fission in small groups, Journal of Anthropological Research 33, 452-473 (1977).
- CElegans*: 306 and 2345 edges. A directed, weighted network representing the neural network of C. Elegans. Data compiled by D. Watts and S. Strogatz. Reference: D. J. Watts and S. H. Strogatz, Nature 393, 440-442 (1998). Original experimental data taken from J.

[1]http://wiki.gephi.org/index.php/Datasets

G. White, E. Southgate, J. N. Thompson, and S. Brenner, Phil. Trans. R. Soc. London 314, 1-340 (1986).

- Diseasome*: 1419 nodes and 3926 edges. A network of disorders and disease genes linked by known disordergene associations, indicating the common genetic origin of many diseases. Reference: The Human Disease Network, Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabsi A-L (2007), Proc Natl Acad Sci USA 104:8685-8690
- Random100: A random network of 100 nodes and 246 edges.
- Random250: A random network of 250 nodes and 1538 edges.
- ScaleFree100: A network of 100 nodes and 150 edges generated with the ”Complex Generators” plugin for Gephi[2]. It uses the ”Barabasi Albert Simplified B” model.
- ScaleFree250: A network of 250 nodes and 1000 edges generated with the ”Complex Generators” plugin for Gephi.
- Internet*: A network of 22963 nodes and 48436 edges. A symmetrized snapshot of the structure of the Internet at the level of autonomous systems, reconstructed from BGP tables posted by the University of Oregon Route Views Project. This snapshot was created by Mark Newman from data for July 22, 2006 and is not previously published.

B. Benchmark table

Dataset	Layout	Ideal convergence iteration	Quality after the most efficient phase	Quality at 750 iterations
SiSWordsFR	YH	47	0.32	0.25
SiSWordsFR	**FA2**	**21**	**0.25**	0.19
SiSWordsFR	**FA2 LinLog**	37	0.26	**0.16**
LesMiserables	YH	46	0.42	0.4
LesMiserables	**FA2**	**41**	**0.39**	0.25
LesMiserables	**FA2 LinLog**	113	0.43	**0.24**
LesMiserables	FR	491	0.41	0.37

[2]http://gephi.org/plugins/complex-generators/

KarateClub	YH	53	0.43	0.41
KarateClub	**FA2**	**51**	**0.35**	0.35
KarateClub	**FA2 LinLog**	236	0.36	**0.31**
KarateClub	FR	431	0.42	0.42
CElegans	YH	**57**	0.51	0.48
CElegans	**FA2**	95	0.44	0.43
CElegans	**FA2 LinLog**	192	**0.42**	**0.4**
CElegans	FR	300	0.45	0.46
Diseasome	YH	97	**0.25**	0.1
Diseasome	**FA2**	**62**	0.44	**0.07**
Diseasome	**FA2 LinLog**	108	0.57	0.08
Diseasome	FR	829	N/A	0.24
Random100	YH	82	**0.54**	0.55
Random100	**FA2**	**70**	0.55	0.51
Random100	**FA2 LinLog**	285	0.58	**0.49**
Random100	FR	588	**0.54**	0.54
Random250	YH	122	0.77	0.77
Random250	**FA2**	**114**	0.72	0.71
Random250	**FA2 LinLog**	432	**0.71**	**0.69**
Random250	FR	454	0.74	0.73
ScaleFree100	YH	**48**	0.45	0.43
ScaleFree100	**FA2**	65	**0.44**	**0.34**
ScaleFree100	**FA2 LinLog**	252	0.45	0.37
ScaleFree100	FR	723	0.46	0.46
ScaleFree250	YH	97	0.68	0.68
ScaleFree250	**FA2**	**84**	**0.63**	0.62
ScaleFree250	**FA2 LinLog**	293	0.64	**0.58**
ScaleFree250	FR	453	0.67	0.67

Internet	YH	**63**	**0.36**	0.27
Internet	**FA2**	83	0.43	0.38
Internet	**FA2 LinLog**	441	0.38	**0.26**