

Mapping new models of complex data

R. Lambiotte

Department of Mathematics

Research group on Complexity and Networks

Namur Center for Complex Systems

University of Namur, Belgium



Computational social science

Small-scale questionnaire-based approaches



Fingerprints of individuals in electronic media (offline: mobile phone, or online: email, Facebook, etc.)

+ Large-scale experiments in online media

Possibility to analyse the dynamics and organisation of large-scale social systems



Social scientists (their interactions)

Psychologist
(people)

Urban planners
(where they live)



Social scientists
(their interactions)

Psychologist
(people)

Urban planners
(where they live)



Physicists
(models)

Computer scientist
(data)

Applied mathematician
(algorithms)

Social scientists (their interactions)

Psychologist
(people)

Urban planners
(where they live)

M. Szell (MIT) and S. Thurner (Vienna) => Online games

E. Fleury (ENS Lyon), A. Friggeri (Facebook), and D. Quercia and M. Kosinski (Cambridge) => Personality

Mason Porter and Till Hoffman (Oxford) => Modeling of temporal networks

Tim Evans (Imperial College) and Pietro Panzarasa (Queen Mary) => Communities

Michael Gastner (Bristol) => Mobility

Vsevolod Salnikov (UNamur) => App development, and networks

Lionel Tabourier (UNamur) and JC Delvenne (Louvain) => Algorithms for temporal nets

Martin Rosvall (Umea) => Communities and ranking

C. Mascolo, A. Noulas (Cambridge) and S. Scellato (Google) => Mobility

Physicists
(models)

Computer scientist
(data)

Applied mathematician
(algorithms)

Social scientists
(their interactions)

Urban planners
(where they live)

Psychologist
(people)



- Physics
- Applied Mathematics
- Mathematics
- Neuroscience
- Bioengineering

Physicists
(models)

Computer scientist
(data)

Applied mathematician
(algorithms)

Social scientists
(their interactions)

10^2

Psychologist
(people)

10^0

Urban planners
(where they live)

10^4-10^6

Physicists
(models)

∞

Applied mathematician
(algorithms)

10^6-10^9

Computer scientist
(data)

10^6-10^9

Social scientists
(their interactions)

10^2

Psychologist
(people)

10^0

Urban planners
(where they live)

10^4-10^6

Different scales
Different methodologies
Different interests

Yet the same problem



Bridge the gap
Enhance the dialogue between the fields

Physicists
(models)

∞

Computer scientist
(data)

10^6-10^9

10^6-10^9

Applied mathematician
(algorithms)

Plethora of new services - new opportunities

Interweaving the social and the physical world

Offering more and more refined data about individuals and their interactions



New data need new methods and new models

New models?

Call Data Records Database

Cell site location		Cell site location								
Cell ID (A)		Cell ID (B)								
Start cell	End cell	Start cell	End cell	Number making call (A)	Number receiving call (B)	Start date/time	Call duration (min:sec)	Call type	CDR type	Dialled digits
'0301063492	'	'	'	447875477828	'	03/12/2010 20:40	01:12	PS	GPRS	'
'0014451843	'	'	'	447875477828	'	03/12/2010 20:41	00:14	PS	GPRS	'
'0301052339	'	'	'	447875477828	'	03/12/2010 20:41	51:34:00	PS	GPRS	'
'	'	'0301052339	'0301052339	447875477828	447772000987	03/12/2010 20:54	00:12	CS	MTC	'
'0301037492	'0301037492	'	'	447875477828	7772000987	03/12/2010 20:54	00:11	CS	MOC	447772000987
'0014457988	'	'	'	447875477828	'	03/12/2010 21:33	04:47	PS	GPRS	'
'0301052339	'	'	'	447875477828	'	03/12/2010 21:38	21:51	PS	GPRS	'
'0301052339	'	'	'	447875477828	'	03/12/2010 22:00	60:00:00	PS	GPRS	'
'0301052339	'	'	'	447875477828	'	03/12/2010 23:00	60:00:00	PS	GPRS	'
'	'	'	'	'7708026012	7875477828	03/12/2010 23:52	00:02	CS	PSTNO	07875477828
'6553565535	'6553565535	'	'	447875477828	1447875477828	03/12/2010 23:52	00:02	CS	CF	'
'	'	'	'	7875477828	1447875477828	03/12/2010 23:52	00:02	CS	CTR	'
'	'	'0301037492	'	'447708026012	447875477828	03/12/2010 23:52		SMS	SMST	'



Location,
mobility, etc.



Social relations



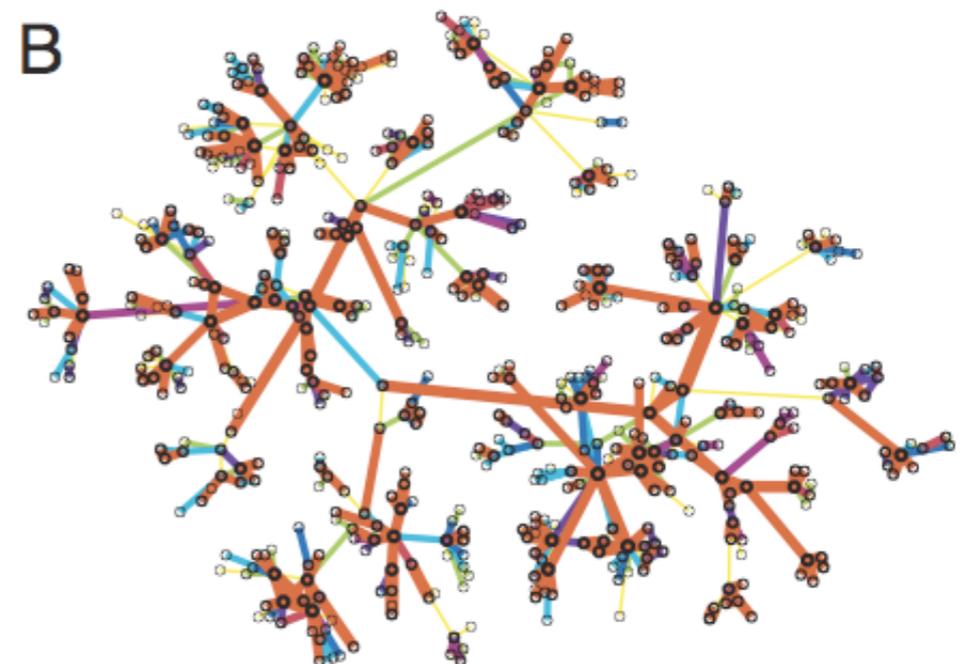
Timings

New models?

Model: complex data are seen through the prism of an abstract model, which allows for the use of formal methods, and an interpretation of the system

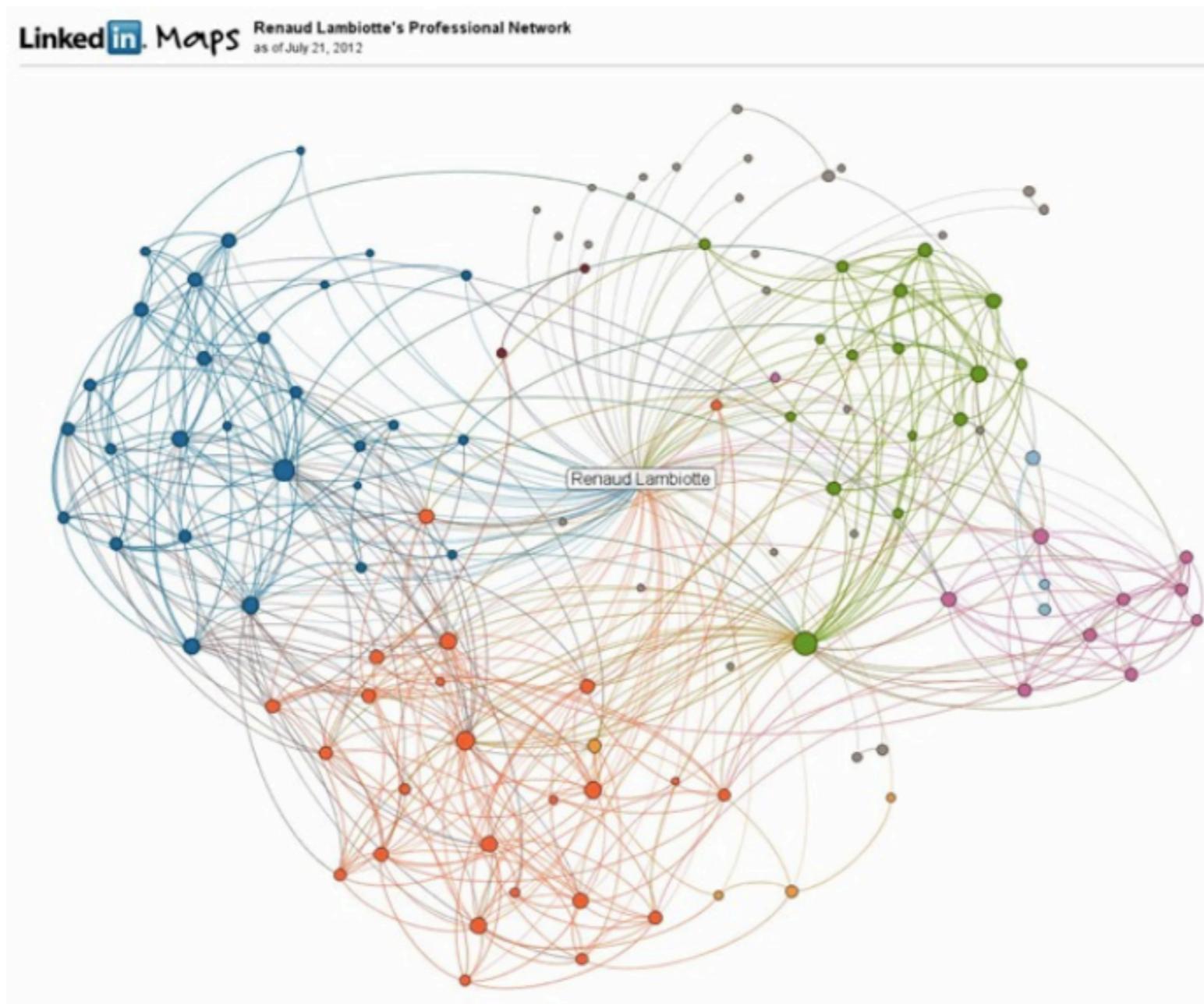
Cell site location		Cell site location								
Cell ID (A)		Cell ID (B)								
Start cell	End cell	Start cell	End cell	Number making call (A)	Number receiving call (B)	Start date/time	Call duration (min:sec)	Call type	CDR type	Dialled digits
'0301063492	'	'	'	447875477828	'	03/12/2010 20:40	01:12	PS	GPRS	'
'0014451843	'	'	'	447875477828	'	03/12/2010 20:41	00:14	PS	GPRS	'
'0301052339	'	'	'	447875477828	'	03/12/2010 20:41	51:34:00	PS	GPRS	'
'	'	'0301052339	'0301052339	447875477828	447772000987	03/12/2010 20:54	00:12	CS	MTC	'
'0301037492	'0301037492	'	'	447875477828	7772000987	03/12/2010 20:54	00:11	CS	MOC	447772000987
'0014457988	'	'	'	447875477828	'	03/12/2010 21:33	04:47	PS	GPRS	'
'0301052339	'	'	'	447875477828	'	03/12/2010 21:38	21:51	PS	GPRS	'
'0301052339	'	'	'	447875477828	'	03/12/2010 22:00	60:00:00	PS	GPRS	'
'0301052339	'	'	'	447875477828	'	03/12/2010 23:00	60:00:00	PS	GPRS	'
'	'	'	'	'7708026012	7875477828	03/12/2010 23:52	00:02	CS	PSTNO	07875477828
'6553565535	'6553565535	'	'	447875477828	1447875477828	03/12/2010 23:52	00:02	CS	CF	'
'	'	'	'	7875477828	1447875477828	03/12/2010 23:52	00:02	CS	CTR	'
'	'	'0301037492	'	'447708026012	447875477828	03/12/2010 23:52		SMS	SMST	'

Network paradigm:
nodes connected by
weighted edges



Network toolbox

Algorithms for network visualisation, community detection, centrality, role detection, etc.



Network science in a nutshell

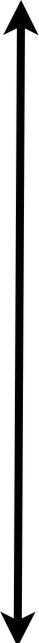
Say that one wants to model epidemic spreading

a) Construction of a network from empirical data, e.g. airline transportation network

b) Definition of a model for epidemic spreading, e.g. meta-population model

 Effect of topology on spreading

Modeling step



Real-world trajectories of the users are not considered: all that matters are fluxes between node (airports)

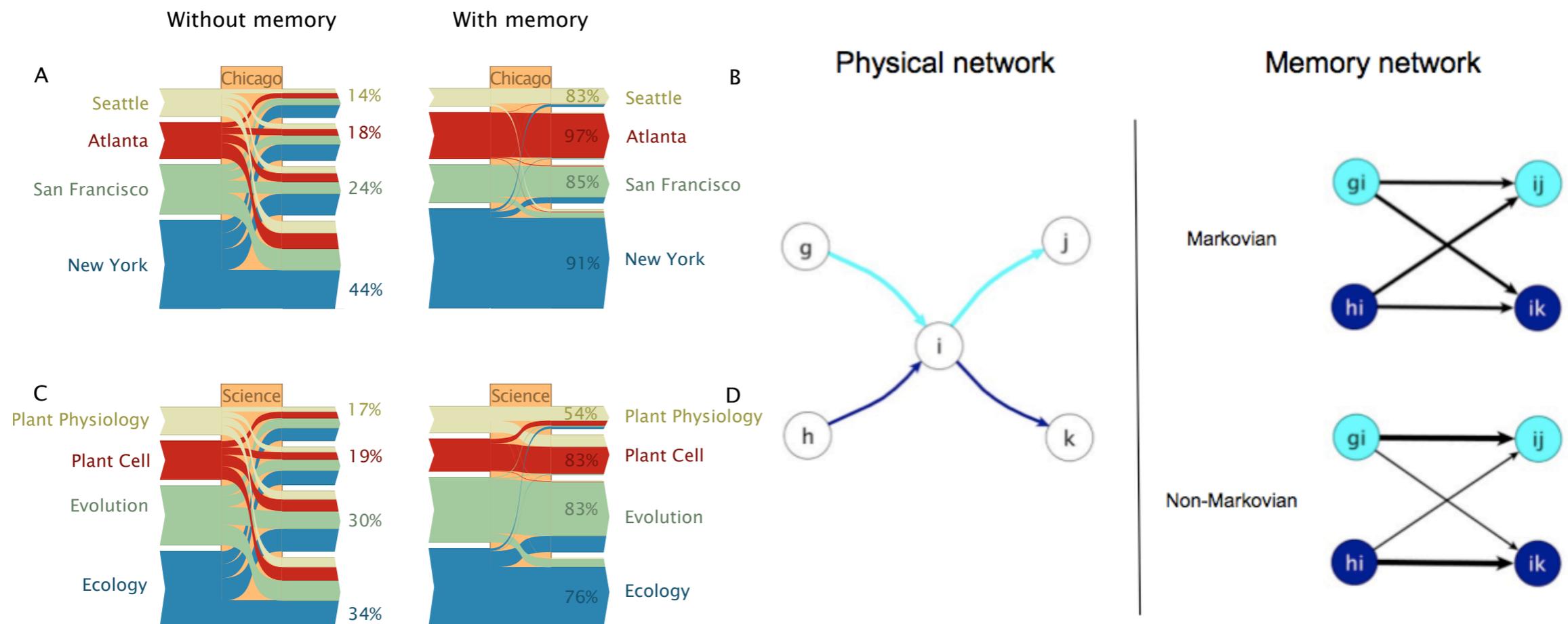
Exact timings at which users meet are not considered: all that matters are fluxes between node (airports)

Now that such data is available: Accuracy of the network paradigm? Other paradigms?

Networks with memory

Second-order Markov: transitions from directed edges to directed edge (memory node)

Mathematics of pathways instead of edges



Networks with Memory, Martin Rosvall, Alcides V. Esquivel, Andrea Lancichinetti, Jevin D. West, Renaud Lambiotte, arXiv: 1305.4807

Slow-Down vs. Speed-Up of Information Diffusion in Non-Markovian Temporal Networks, I Scholtes, N Wider, R Pfitzner, A Garas, C Juan Tesson and F Schweitzer, arXiv:1307.4030

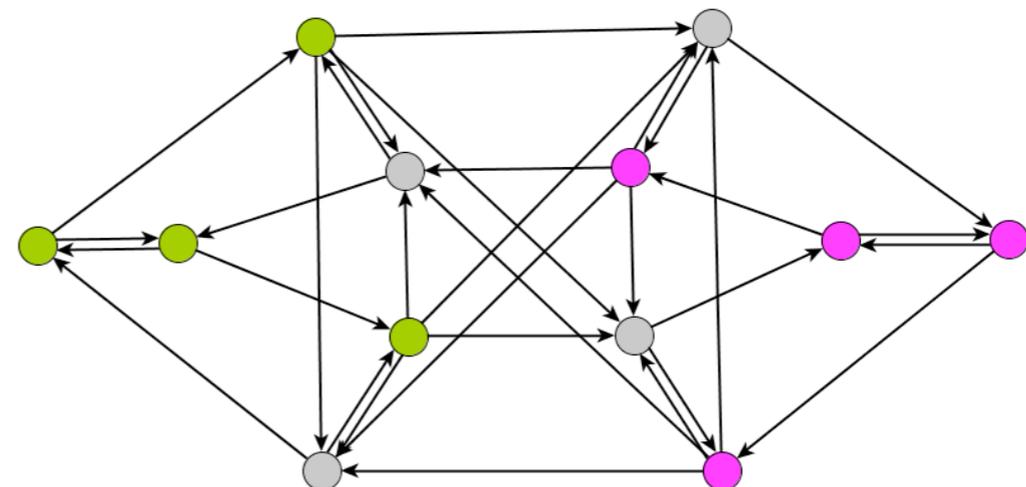
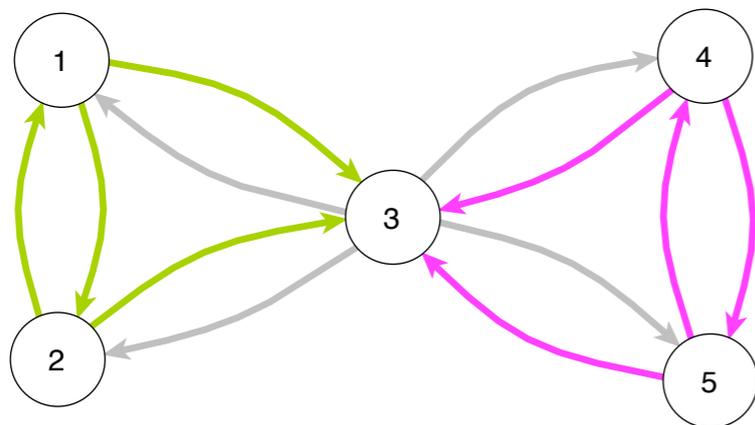
Toolbox for pathways

In several empirical systems: memory constraints on flow are statistically significant and temporal correlations strongly modify flows of probability.

Memory networks are networks, but

- they are much, much larger => computational complexity
- they have a very different structure => require new tools

First insight on community detection, ranking, spreading, but there is a need for appropriate visualisation, statistical and algorithmic tools.



Financial Support

IAP (Belspo): Applied Mathematics

Postdoc opening

FNRS: Temporal networks

Lectureship

ESF COST: KnowEscape

Visualization of the Dutch Research Landscape

KNOWESCAPE

Analyzing the dynamics of information and knowledge landscapes

Home About » Working groups » STSM Events » Publications Private Area » Contact

Design vs. Emergence: visualization of knowledge orders

© Almila Akdag, Cheng Gao, Krzysztof Suchecki, Andrea Scharnhorst

CATEGORY DISTRIBUTION OF WIKIPEDIA & UDC

This donut chart shows the distribution of the nine UDC categories (inner ring) and the 43 top Wikipedia categories (outer ring). Wikipedia categories are further assigned to corresponding UDC classes, and colored accordingly. About 72% of UDC categories belong to Sciences (22% Natural Sciences and 50% Applied Sciences), Wikipedia's 43 top categories, however, are distributed in a much more balanced fashion, with many categories found under Arts, Entertainment & Sports, followed by Science, Knowledge, Organization, and Social Sciences. A large part, 17%, of Wikipedia categories are tagged as 'ambiguous'.

In this version of KnowEscape, our data stems from a 2018 Master Reference File (MRF) of UDC, and has a total of 88,548 classes organized in a tree layout branching out to the depth of nine, organized under nine main classes. The visualization shown here covers the whole UDC network except the auxiliaries, which consists of 55,304 nodes and 55,303 edges. Initial network layout was calculated using DL, IvisOrb in Sci2, the final layout was rendered using the Fruchterman-Reingold layout in Gephi.