machine learning in climate science

Deep graphs – A general framework to represent and analyze heterogenous complex systems across scales

Julia Hellmig

Journal Club

08 June 2021

Chaos

Deep graphs—A general framework to represent and analyze heterogeneous complex systems across scales

Cite as: Chaos **26**, 065303 (2016); https://doi.org/10.1063/1.4952963
Submitted: 16 February 2016 . Accepted: 14 April 2016 . Published Online: 02 June 2016

Dominik Traxl, Niklas Boers, and Jürgen Kurths

View Online   Export Citation   CrossMark

Introduction and Motivation

1

- Current network representations
- Goal and aim of Deep graphs

Deep Graphs Framework

2

- The basics
- Partitioning

Construction of a Deep Graph

3

- Creating Nodes
- Creating Edges
- Creating Supernodes

Application to Precipitation Data

4

- Creation of the Deep Graph
- Results

Current network representations

- Weighted Graphs
- Node weighted networks
- Hypergraphs
- Multilayer networks

But: some features are still missing and no network representation can do all these things

Missing Features

- Interactions between groups on different scales
- Association of information with heterogenous types of objects and relations
-

Aims of Deep Graphs

- Entail and unify existing network representations
- Generalize and extend network representations by the missing features

Benefits of Deep Graphs

- Combine heterogenous datasets
- Integrate a priori knowledge of groups of objects and their relations
- Conduct analysis of interrelations of the system within the network representation

Introduction and Motivation

1

- Current network representations
- Goal and aim of Deep graphs

Deep Graphs Framework

2

- The basics
- Partitioning

Construction of a Deep Graph

3

- Creating Nodes
- Creating Edges
- Creating Supernodes

Application to Precipitation Data

4

- Creation of the Deep Graph
- Results

Deep graphs – The basis

- A finite, directed graph
- Every node can contain features
- Every edge can contain relations
- Every feature and relation is of a type

$$G = (V, E), \tag{1}$$

where V is a set of $n := |V|$ nodes,

$$V = \{V_i \,|\, i \in \{1, 2, ..., n\}\}, \tag{2}$$

and E is a set of $m := |E|$ directed edges, given by

$$E \subseteq \{E_{ij} \,|\, i, j \in \{1, 2, ..., n\}\} =: E'. \tag{3}$$

Deep graphs - A general framework to represent and analyze heterogenous complex systems across scales

Julia Hellmig

Deep graphs

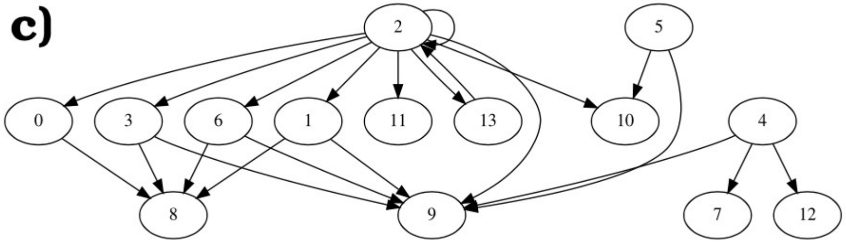
a)

i	category	name	occupation	tax	age	income	gender	political ideology
0	person	Marge Gunderson	state police officer	n/a	46	50	female	egalitarianism
1	person	Frank Underwood	president	n/a	21	1200	male	conservatism
2	person	Avon Barksdale	chief executive officer	n/a	91	23000	male	anarchism
3	person	Leah Gould	judge	n/a	24	85	female	egalitarianism
4	person	Valeria Velez	chief editor	n/a	33	45	female	environmentalism
5	person	Mr. Garrison	teacher	n/a	29	35	male	environmentalism
6	person	Rust Cohle	federal police officer	n/a	61	40	male	egalitarianism
7	newspaper	Twin Peaks Gazette	n/a	n/a	198	n/a	n/a	conservatism
8	state	Krakozhia	n/a	state tax, 10%	198	n/a	n/a	conservatism
9	state	Tomania	n/a	state tax, 5%	141	n/a	n/a	conservatism
10	party	Ingsoc	n/a	n/a	198	n/a	n/a	conservatism
11	company	Viktor's Gun Shop	n/a	n/a	141	n/a	n/a	egalitarianism
12	think tank	Policy Solutions Association	n/a	n/a	198	n/a	n/a	conservatism
13	bank	Commonwealth Shared Risk	n/a	n/a	141	n/a	n/a	neoliberalism

b)

i	j	lives in	works for	monthly cash flow	kind of payment
0	8	TRUE	TRUE	5	tax
1	8	TRUE	TRUE	120	tax
1	9	n/a	TRUE	0	tax
2	0	n/a	n/a	60	bribe
2	1	n/a	n/a	500	bribe
2	2	n/a	TRUE	n/a	n/a
2	3	n/a	n/a	90	bribe
2	6	n/a	n/a	70	bribe
2	9	TRUE	n/a	20	tax
2	10	n/a	n/a	250	donation
2	11	n/a	n/a	20	expense
2	13	n/a	n/a	10000	investment
3	8	TRUE	TRUE	9	tax
3	9	n/a	TRUE	0	tax
4	7	n/a	TRUE	n/a	n/a
4	9	TRUE	n/a	3	tax
4	12	n/a	TRUE	5	donation
5	9	TRUE	TRUE	1	tax
5	10	n/a	TRUE	1	donation
6	8	TRUE	TRUE	4	tax
6	9	n/a	TRUE	0	tax
13	2	n/a	n/a	5000	investment

c)



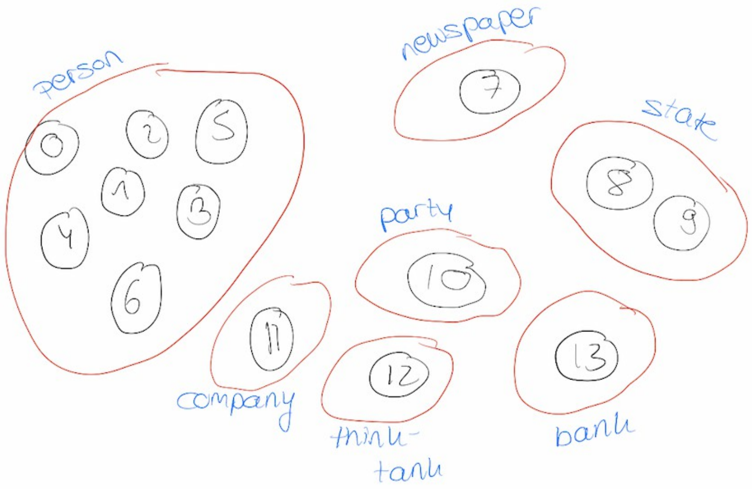
Julia Hellmig

Supernodes and Superedges

- Any subset of nodes of a graph can be grouped into a **supernode**
- Edges can be placed between supernodes and between supernodes and nodes
- Any subset of edges of a graph can be grouped into a **superedge**
- A **supergraph** results by partitioning the node set and subsequently partitioning the edge set

Partitioning of the graph by category

Results in seven supernodes



Intersection Partitions

- Choose partitions of a graph of which a intersection partition should be created

- All nodes that belong to all partitions (supernodes) of the chosen ones belong to the intersection partition

- An edge belongs to the intersection partition if it originates and targets a node in a supernode of the intersection partition

- Find potentially informative partitions of a graph

  ϴ all edges originating from nodes with a political ideology of "egalitarism" target nodes of the category "state"

- Compute similarity measures between different partitions

  ϴ calculate the normalized variation of information metric, Jaccard index or the normalized information index

Introduction and Motivation

1

- Current network representations
- Goal and aim of Deep graphs

Deep Graphs Framework

2

- The basics
- Partitioning

Construction of a Deep Graph

3

- Creating Nodes
- Creating Edges
- Creating Supernodes

Application to Precipitation Data

4

- Creation of the Deep Graph
- Results

1. Every row in a Dataframe is a Node

2. Columns are features of the nodes

3. Create edges

   a) Connectors: create edges between nodes with relations
   b) Apply a priori knowledge about The relations of the nodes
   c) Selectors: filter the edges based on selective properties

4. Create a supergraph by identifying a partition of the graph

5. Compute partition-specific features to the supernodes

	latitude	longitude	time	t2m	x	y	day	month	year	daily_mag	g_id	itime	cp
140	53.939999	6.72	2010-04-07 17:00:00	10.138086	1	0	7	4	2010	1.371126	9	0	255
285	53.939999	8.22	2010-04-07 17:00:00	10.917261	2	0	7	4	2010	0.971329	14	0	255
8389	49.439999	11.22	2010-04-29 17:00:00	24.636042	4	3	29	4	2010	1.747910	21	22	171
4187	50.939999	5.22	2010-04-29 17:00:00	23.670282	0	2	29	4	2010	2.429068	2	22	171
7492	49.439999	8.22	2010-04-29 17:00:00	24.687281	2	3	29	4	2010	2.017557	11	22	171
...	...	...	...	...	...	...	...	...	...	...	...	...	...
12188	47.939999	14.22	2020-12-06 17:00:00	6.121362	6	4	6	12	2020	1.726787	30	3896	186
139	53.939999	5.22	2020-12-14 17:00:00	10.407801	0	0	14	12	2020	1.455739	4	3904	0
12189	47.939999	14.22	2020-12-22 17:00:00	7.256280	6	4	22	12	2020	1.092964	30	3912	215
12190	47.939999	14.22	2020-12-23 17:00:00	8.554590	6	4	23	12	2020	3.267378	30	3913	215

	n_nodes	time_amin	time_amax	itime_amin	itime_amax	t2m_amax	HWMId_magnitude	latitude_mean	longitude_mean	g_ids	n_unique_g_ids	dt
cp												
1	593	2018-07-15 17:00:00	2018-08-09 17:00:00	3021	3046	36.020164	976.370285	50.907115	9.763002	{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,...	35	25 days
2	351	2013-07-15 17:00:00	2013-07-30 17:00:00	1195	1210	35.480736	426.771580	50.452819	9.878120	{0, 1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 15	32	15 days

Introduction and Motivation

1

- Current network representations
- Goal and aim of Deep graphs

Deep Graphs Framework

2

- The basics
- Partitioning

Construction of a Deep Graph

3

- Creating Nodes
- Creating Edges
- Creating Supernodes

Application to Precipitation Data

4

- Creation of the Deep Graph
- Results

Application to global precipitation data

- Precipitation measurements from 1998 to 2014
- Local formations of spatio-temporal clusters of extreme precipitation events

$$\boldsymbol{G = (V, E)}$$

a) $\boldsymbol{V_i}$

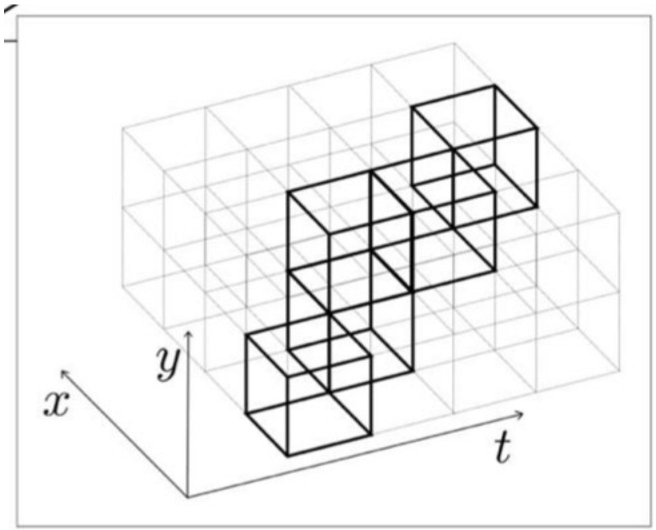
feature	symbol	type of feature	given by
F_i^1	L_i	geographical label	$(x_i, y_i) \leftrightarrow L_i$
F_i^2	$\underline{x}_i$	space-time coordinates	$(lon_i, lat_i, t_i) \leftrightarrow (x_i, y_i, t_i)$
F_i^3	a_i	surface area	$(111\text{km})^2 \cdot (0.25)^2 \cdot \cos\left(\frac{2\pi}{360°} \cdot lat_i\right)$
F_i^4	r_i	precipitation rate	given
F_i^5	v_i	vol. of water precipitated	$a_i \cdot r_i \cdot 3\text{h}$
F_i^6	C_i	cluster membership	connceted components
F_i^7	F_i	family membership	linkage clustering

b) $\boldsymbol{E_{ij}}$

relations	condition	symbol	type of relation	given by
R_{ij}^1	if $\lvert d\alpha_{ij}\rvert \leq 1\ \forall\alpha \in \{x, y, t\}$	$d\underline{x}_{ij}$	spatio-temporal distance	$\underline{x}_j - \underline{x}_i$
$\emptyset$	else			

Create clusters

- Every grid cell in a 3D grid can have 26 possible neighbors
- Create edges between those neighbor nodes based on their spatio-temporal distance
- Clusters of connected nodes are created



Traxl, D., N. Boers, A. Rheinwalt, B. Goswami, and J. Kurths (2016), The size distribution of spatiotemporal extreme rainfall clusters around the globe, Geophys. Res. Lett., 43, 9939–9947, doi:10.1002/2016GL070692.

$$G = (V, E)$$

a) V_i

feature	symbol	type of feature	given by
F_i^1	L_i	geographical label	$(x_i, y_i) \leftrightarrow L_i$
F_i^2	$\underline{x}_i$	space-time coordinates	$(lon_i, lat_i, t_i) \leftrightarrow (x_i, y_i, t_i)$
F_i^3	a_i	surface area	$(111\text{km})^2 \cdot (0.25)^2 \cdot \cos\left(\frac{2\pi}{360°} \cdot lat_i\right)$
F_i^4	r_i	precipitation rate	given
F_i^5	v_i	vol. of water precipitated	$a_i \cdot r_i \cdot 3\text{h}$
F_i^6	C_i	cluster membership	connceted components
F_i^7	F_i	family membership	linkage clustering

b) E_{ij}

relations	condition	symbol	type of relation	given by
R_{ij}^1	if $\lvert d\alpha_{ij}\rvert \leq 1 \; \forall \alpha \in \{x, y, t\}$	$d\underline{x}_{ij}$	spatio-temporal distance	$\underline{x}_j - \underline{x}_i$
$\emptyset$	else			

Partitioning

- Partitioning of the graph by clusters
- Creation of spatio-temporal extreme rainfall clusters
- Computation of supernode features
- Creation of superedges between clusters that have a strong regional overlap
- Agglomerative, hierarchial clustering of the clusters into familes

$$\boldsymbol{G^C = (V^C, E^C)}$$

a) $\boldsymbol{V_i^C}$

feature	symbol	type of feature	given by
${}^C F_i^1$	t_i^{min}	starting time	$\min_{j \in S} t_j$
${}^C F_i^2$	t_i^{max}	end time	$\max_{j \in S} t_j$
${}^C F_i^3$	Δt_i	time span	$t_i^{max} - t_i^{min}$
${}^C F_i^4$	v_i^{sum}	total vol. of water precipitated	$\sum_{j \in S} v_j$
${}^C F_i^5$	L_i^{set}	set of geographical labels	$\{L_j \mid j \in S\}$
${}^C F_i^6$	a_i^{sum}	spatial coverage	$\sum_{L_j \in L_i^{set}} A(L_j)$
${}^C F_i^7$	F_i	family membership	linkage clustering

where $S = \{j \mid j \in \{1, 2, ..., n\} \wedge p^C(V_j) = C_i\}$

b) $\boldsymbol{E_{ij}^C}$

relation	symbol	type of relation	given by
${}^C R_{ij}^1$	dt_{ij}	temporal distance between clusters	$t_j^{min} - t_i^{min}$
${}^C R_{ij}^2$	IC_{ij}	intersection cardinality	$\lvert L_i^{set} \cap L_j^{set} \rvert$
${}^C R_{ij}^3$	IS_{ij}	intersection strength	$\frac{\mathrm{IC}_{ij}}{\min\{\lvert L_i^{set}\rvert, \lvert L_j^{set}\rvert\}}$

- Partitioning of the graph based on family membership of the nodes

  ɵ now extensive statistical analysis could be performed

$$G^F = (V^F, E^F)$$

a) V_i^F

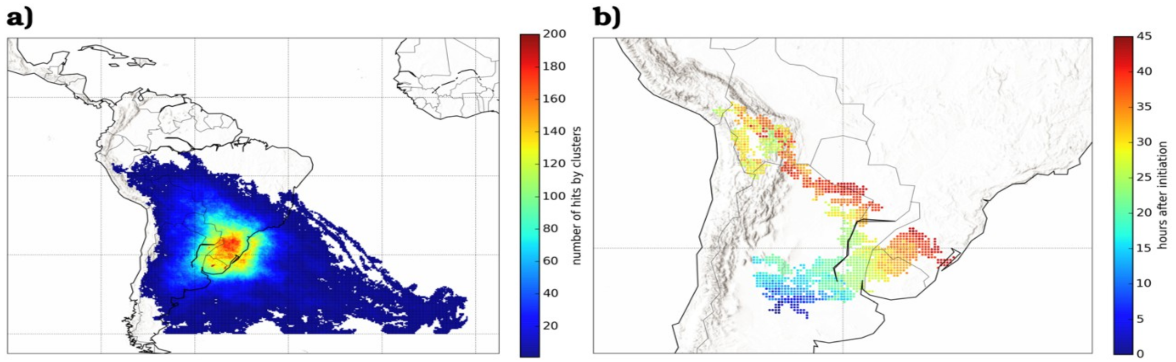
feature	symbol	type of feature	given by
${}^{F}F_i^1$	T_i^{min}	tuple of start times	$(t_j^{min})_{j\in S}$
${}^{F}F_i^2$	T_i^{max}	tuple of end times	$(t_j^{max})_{j\in S}$
${}^{F}F_i^3$	ΔT_i	tuple of time spans	$(\Delta t_j)_{j\in S}$
${}^{F}F_i^4$	${}^{F}v_i^{sum}$	total vol. of water precipitated	$\sum_{j\in S} v_j^{sum}$
${}^{F}F_i^5$	${}^{F}L_i^{set}$	set of geographical locations	$\bigcup_{j\in S} L_j^{set}$
${}^{F}F_i^6$	${}^{F}a_i^{sum}$	spatial coverage	$\sum_{L_j\in {}^{F}L_i^{set}} A(L_j)$
where $S = \{j \,\vert\, j \in \{1,2,...,n^c\} \wedge p^F(V_j^C) = F_i\}$			

b) E_{ij}^F

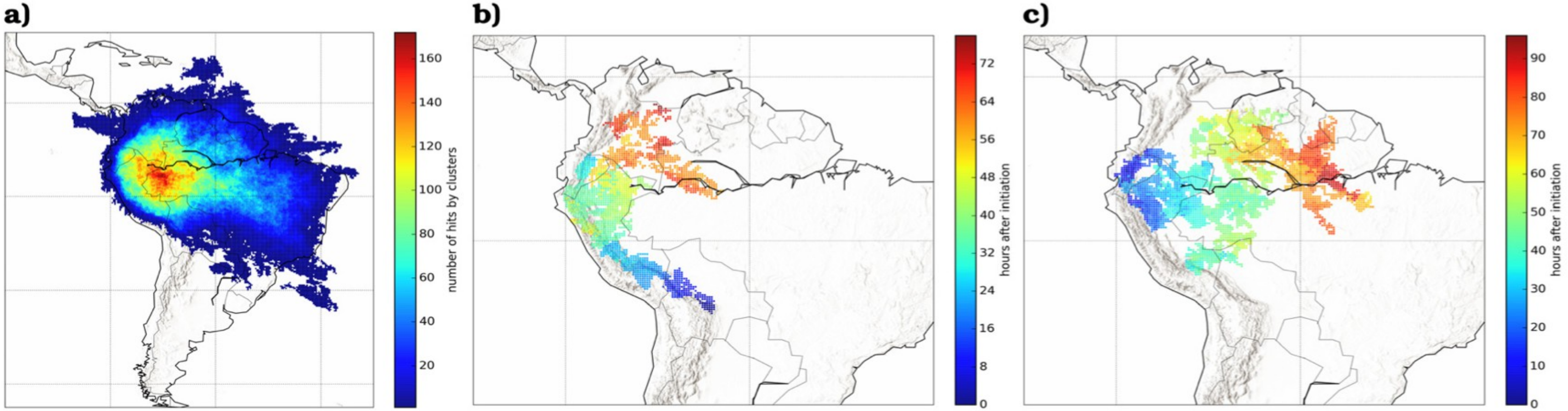
relation	symbol	type of relation	given by
${}^{F}R_{ij}^1$	$(\mathrm{dt}_{uv})_{(u,v)\in S}$	tuple of inter- or intra-family temporal distances	E^C
${}^{F}R_{ij}^2$	$(\mathrm{IC}_{uv})_{(u,v)\in S}$	tuple of inter- or intra-family intersection cardinalities	E^C
${}^{F}R_{ij}^3$	$(\mathrm{IS}_{uv})_{(u,v)\in S}$	tuple of inter- or intra-family intersection strengths	E^C
where $S = \{(u,v) \,\vert\, u,v \in \{1,2,...,n^c\} \wedge p^F(V_u^C) = F_i \wedge p^F(V_v^C) = F_j\}$			

Visualization of the time evolution of extreme rainfall clusters

- Concentration on two families of extreme precipitation clusters in South America



- Concentration on two families of extreme precipitation clusters in South America



Conclusion

- Introduction of a framework to describe and analyze heterogenous multiscale systems
- Based on network theory
- Includes an extensive software package that is easy to handle and apply to own data

Sources

- Dominik Traxl, Niklas Boers and Jürgen Kurths, “Deep graphs-A general framework to represent and analyze heterogenous complex systems across scales”, Chaos: An interdisciplinary Journal of Nonlinear Science 26, 065303 (2016) https://doi.org10.1063/1.4952963
- Traxl, D., N. Boers, A. Rheinwalt, B. Goswami, and J. Kurths (2016), The size distribution of spatiotemporal extreme rainfall clusters around the globe, Geophys. Res. Lett., 43, 9939–9947, doi:10.1002/2016GL070692.