# Scalable Partitioning of Large Complex Networks

Luce le Gorrec

UK IC Postodoctoral Research Fellow

Mathematics and Statistics Department,

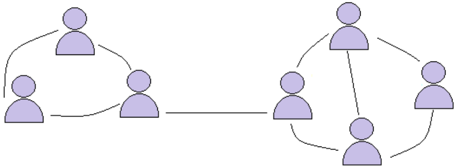
University of Strathclyde,

Glasgow, United Kingdom

Royal Academy of Engineering

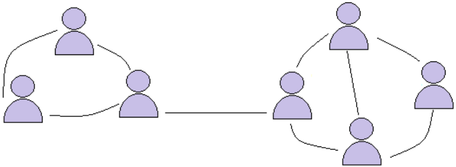
# Graph Partitioning

**Graph** (or network): entities (nodes) connected by relations (edges).



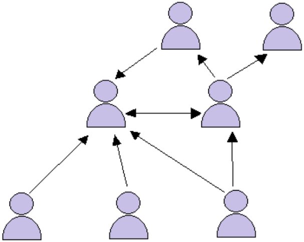
# Graph Partitioning

**Graph** (or network): entities (nodes) connected by relations (edges).
Different kinds of graphs: directed, weighted, evolving...



# Graph Partitioning

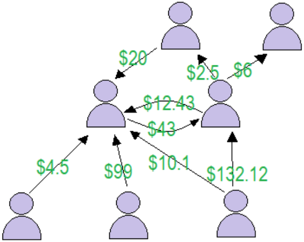
**Graph** (or network): entities (nodes) connected by relations (edges).
Different kinds of graphs: **directed**, weighted, evolving...



E.g. Who-follows-who network.

# Graph Partitioning

**Graph** (or network): entities (nodes) connected by relations (edges).
Different kinds of graphs: directed, **weighted**, evolving...

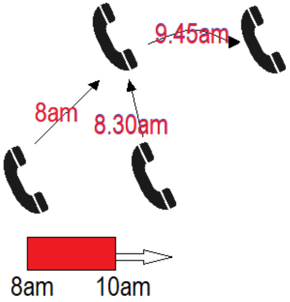



E.g. Money transfer network.

Royal Academy of Engineering

# Graph Partitioning

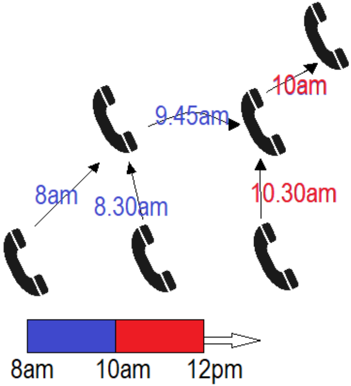
**Graph** (or network): entities (nodes) connected by relations (edges).
Different kinds of graphs: directed, weighted, **evolving**...




Royal Academy of Engineering

# Graph Partitioning

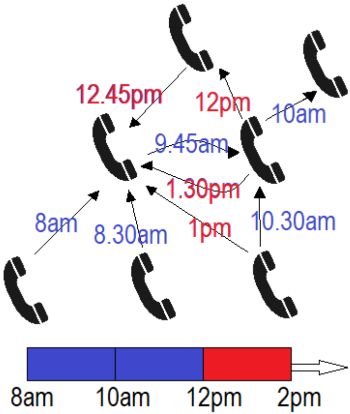
**Graph** (or network): entities (nodes) connected by relations (edges).
Different kinds of graphs: directed, weighted, **evolving**...




Royal Academy of Engineering

# Graph Partitioning

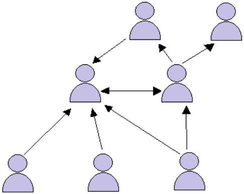
**Graph** (or network): entities (nodes) connected by relations (edges).
Different kinds of graphs: directed, weighted, **evolving**...




Royal Academy of Engineering

# Graph Partitioning

**Graph** (or network): entities (nodes) connected by relations (edges).
Different kinds of graphs: directed, weighted, evolving...



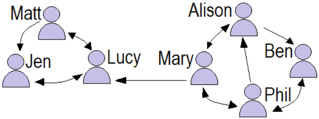
$\Longrightarrow$ **Our study case : directed, unweighted, static.**

Royal Academy of Engineering

# Graph Partitioning

What : A partition of nodes with high (low) density of edges within (between) the groups: a **community structure**.

A Network




Its Adjacency Matrix

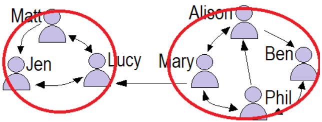
| | Alison | Ben | Jen | Lucy | Mary | Matt | Phil |
|---|---|---|---|---|---|---|---|
| **Alison** | ♪ | x | | | x | | |
| **Ben** | | ♪ | | | | | x |
| **Jen** | | | ♪ | x | | | |
| **Lucy** | | | x | ♪ | | x | |
| **Mary** | x | | | x | ♪ | | x |
| **Matt** | | | x | x | | ♪ | |
| **Phil** | x | x | | | x | | ♪ |

Royal Academy of Engineering

# Graph Partitioning

What : A partition of nodes with high (low) density of edges within (between) the groups: a **community structure**.

A Network




Its Adjacency Matrix

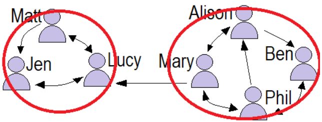
| | Jen | Matt | Lucy | Mary | Alison | Ben | Phil |
|---|---|---|---|---|---|---|---|
| Jen | ♪ | | x | | | | |
| Matt | x | ♪ | x | | | | |
| Lucy | x | x | ♪ | | | | |
| Mary | | | x | ♪ | x | | x |
| Alison | | | | x | ♪ | x | |
| Ben | | | | | | ♪ | x |
| Phil | | | | x | x | x | ♪ |

Royal Academy of Engineering

# Graph Partitioning

What : A partition of nodes with high (low) density of edges within (between) the groups: a **community structure**.

A Network




Its Adjacency Matrix

| | *Jen* | *Matt* | *Lucy* | *Mary* | *Alison* | *Ben* | *Phil* |
|---|---|---|---|---|---|---|---|
| ***Jen*** | ♪ | | x | | | | |
| ***Matt*** | x | ♪ | x | | | | |
| ***Lucy*** | x | x | ♪ | | | | |
| ***Mary*** | | | x | ♪ | x | | x |
| ***Alison*** | | | | x | ♪ | x | |
| ***Ben*** | | | | | | ♪ | x |
| ***Phil*** | | | | x | x | x | ♪ |

Why :

- **Data analysis**: Users spreading terrorist propaganda on Twitter[1].
- **Numerical efficiency**: Analysis of a hundred-million-node network[2].

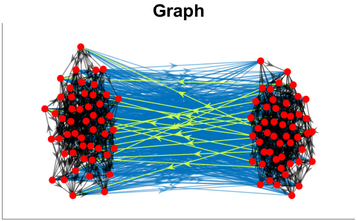
Royal Academy of Engineering

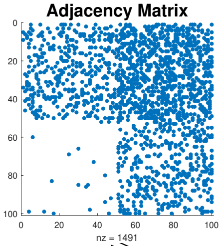
# Graph Partitioning

How: Partitioning a graph: NP-hard:

- But for undirected graphs: Efficient, simple and well-established heuristic (Louvain [3], Metis [4], ...).
- For directed graphs: Nothing as simple or well-established ([5]).

$\Longrightarrow$ "forgetting" edge directions to get undirected networks (suboptimal).



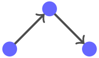

Royal Academy of Engineering

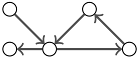
# Motifs

Recent works [6,7,8] focus on **motifs** to partition directed networks.

Motif: A small **induced** subgraph of a certain kind.

*Looking for motifs* $\mathcal{M}$      *in a graph*





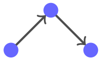
Royal Academy of Engineering

# Motifs

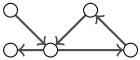
Recent works [6,7,8] focus on **motifs** to partition directed networks.

Motif: A small **induced** subgraph of a certain kind.

*Looking for motifs $\mathcal{M}$*

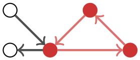


*in a graph*



$\Longrightarrow$

*not a motif $\mathcal{M}$*



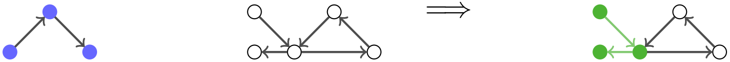
Royal Academy of Engineering

# Motifs

Recent works [6,7,8] focus on **motifs** to partition directed networks.

<u>Motif:</u> A small **induced** subgraph of a certain kind.

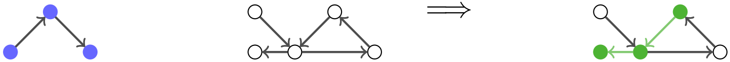
*Looking for motifs* $\mathcal{M}$ *in a graph* *a motif* $\mathcal{M}$



Recent works [6,7,8] focus on **motifs** to partition directed networks.

Motif: A small **induced** subgraph of a certain kind.

*Looking for motifs $\mathcal{M}$* *in a graph* *another motif $\mathcal{M}$*



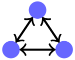
# Motifs

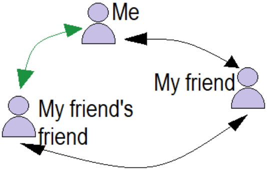
Recent works [6,7,8] focus on **motifs** to partition directed networks.

Motif: A small **induced** subgraph of a certain kind.

**Motifs express complex notions in networks**:

Friends of my friends are my friends:






Royal Academy of Engineering

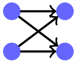
# Motifs

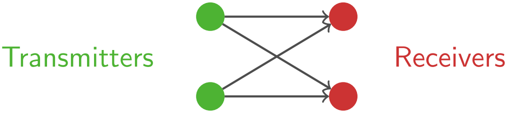
Recent works [6,7,8] focus on **motifs** to partition directed networks.

Motif: A small **induced** subgraph of a certain kind.

**Motifs express complex notions in networks**:

Cooperative propagation of information:

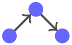
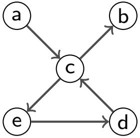




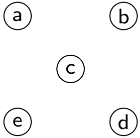

Royal Academy of Engineering

# Motifs

An application of motifs: an undirected network induced from a directed one:

*Looking for motifs*  *in the graph* 


*provides the graph:*



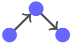

*whose adjacency matrix is:*

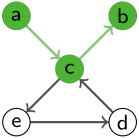
| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | | | | | |
| b | | | | | |
| c | | | | | |
| d | | | | | |
| e | | | | | |

Royal Academy of Engineering

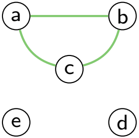
# Motifs

An application of motifs: an undirected network induced from a directed one:

*Looking for motifs*  *in the graph*




*provides the graph:*



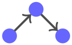
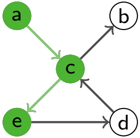

*whose adjacency matrix is:*

| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | | 1 | 1 | | |
| b | 1 | | 1 | | |
| c | 1 | 1 | | | |
| d | | | | | |
| e | | | | | |

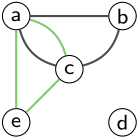
Royal Academy of Engineering

# Motifs

An application of motifs: an undirected network induced from a directed one:

*Looking for motifs*  *in the graph* 


*provides the graph:*

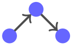
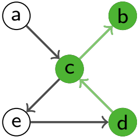



*whose adjacency matrix is:*

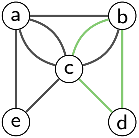
| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | | 1 | 2 | | 1 |
| b | 1 | | 1 | | |
| c | 2 | 1 | | | 1 |
| d | | | | | |
| e | 1 | | 1 | | |

# Motifs

An application of motifs: an undirected network induced from a directed one:

*Looking for motifs*  *in the graph* 


*provides the graph:*



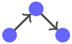

*whose adjacency matrix is:*

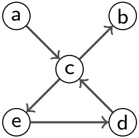
| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | | 1 | 2 | | 1 |
| b | 1 | | 2 | 1 | |
| c | 2 | 2 | | 1 | 1 |
| d | | 1 | 1 | | |
| e | 1 | | 1 | | |

Royal Academy of Engineering

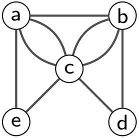
# Motifs

An application of motifs: an undirected network induced from a directed one:

*Looking for motifs*  *in the graph*




*provides the graph:*



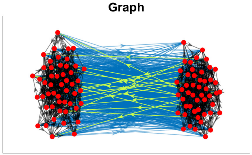

*whose adjacency matrix is:*

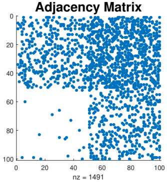
| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | | **1** | **2** | | **1** |
| b | **1** | | **2** | **1** | |
| c | **2** | **2** | | **1** | **1** |
| d | | **1** | **1** | | |
| e | **1** | | **1** | | |

$\Longrightarrow$ the **Benson Graph** of the initial network [6].

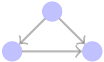
Royal Academy of Engineering

# Motifs and Benson Graphs

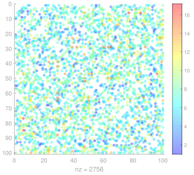


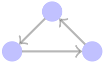

Adjacency Matrix
of the Benson graph
of motif:



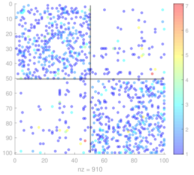
⇓




Adjacency Matrix
of the Benson graph
of motif:

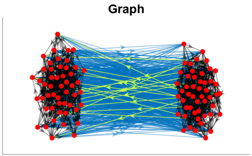


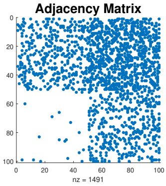
⇓



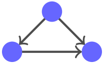

Royal Academy of Engineering

# Motifs and Benson Graphs

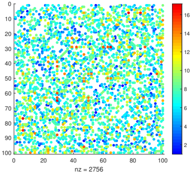


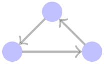

Adjacency Matrix
of the Benson graph
of motif:



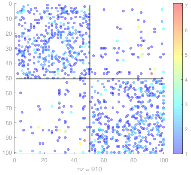
⇓




Adjacency Matrix
of the Benson graph
of motif:

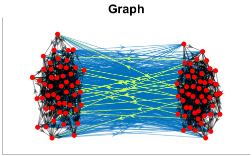


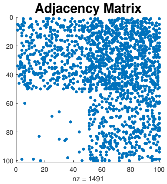
⇓



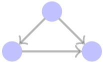

Royal Academy of Engineering

# Motifs and Benson Graphs

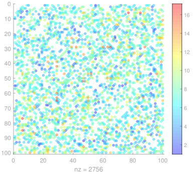


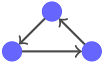

Adjacency Matrix
of the Benson graph
of motif:



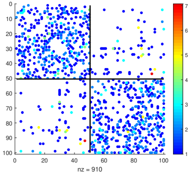
$\Downarrow$




Adjacency Matrix
of the Benson graph
of motif:



$\Downarrow$




Royal Academy of Engineering

# Which motif(s) should we use?

Assessing a motif significance:

- With inferred knowledge.
- ✗ What if no available knowledge?

# Which motif(s) should we use?

Assessing a motif significance:

- With inferred knowledge.
- ✗ What if no available knowledge?
- Statistical significance (ZScore) [9]: *'Significant motifs appear more often than "by chance".'*
  A network $G$, a motif $\mathcal{M}$, a sequence of random networks $\{H_1, ..., H_k\}$ :

$$ZScore(\mathcal{M}) = \frac{\#\text{motifs } \mathcal{M} \text{ in } G - mean(\#\text{motifs } \mathcal{M} \text{ in } H_i)}{(std(\#\text{motifs } \mathcal{M} \text{ in } H_i) + \varepsilon)}.$$

- ✗ Expensive, no consensus about random model.

Royal Academy of Engineering

# Which motif(s) should we use?

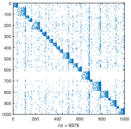
Assessing a motif significance:

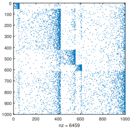
- With inferred knowledge.
- ✗ What if no available knowledge?
- Statistical significance (ZScore) [9]: *'Significant motifs appear more often than "by chance".'*
- ✗ Expensive, no consensus about random model.
- Our proposal : Assess the **discriminatory capacity of motifs on a dataset**: a measure ($\gamma$-score) derived from a feature selection process based on Principal Component Analysis. *(A preprint submitted).*

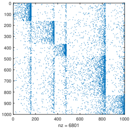
Royal Academy of Engineering

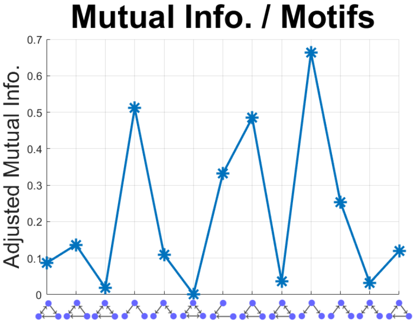
# Which motif(s) should we use? Work in progress

Detecting communities in modular networks [10] using Louvain applied on the Benson graphs.



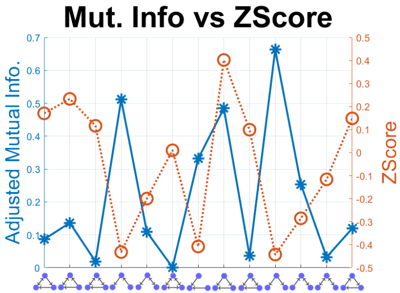



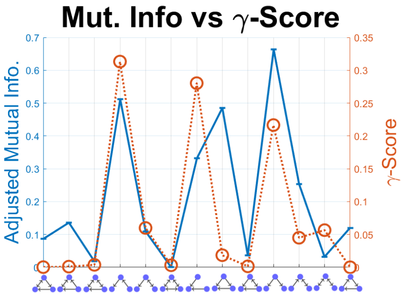




Royal Academy of Engineering

# Which motif(s) should we use? Work in progress

Detecting communities in modular networks [10] using Louvain applied on the Benson graphs.




$\Longrightarrow$ Motifs with highest $\gamma$-score $\sim$ Motifs with highest Adj. Mut. Info.

Royal Academy of Engineering

# Which motif(s) should we use? TO DO

- Expand the preliminary study to confirm/dismiss the correlation high $\gamma$-score/well-detected blocks.
- Focus on local community detection: different motifs may help to detect different communities [11].

# Building the Benson Graphs

- Naive : Whole decomposition of the network (finding all the Benson graphs): FanMod [12].

✗ Prohibitive complexity.

- Some efficient techniques exist [6].

✗ For certain kinds of motifs only.

- We have derived generic formulas to directly compute the Benson adjacency matrices of 3- and 4-node motifs.

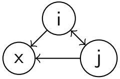
Royal Academy of Engineering

# Building the Benson Graphs: Our formulas

Two observations: A graph $G$ with $n$ nodes, $A$ of dim $n \times n$ its adjacency matrix.

Observation 1: 3 matrices $B, U, N$ of dim $n \times n$ based on $A$ s.t.:

- $B(i,j) = 1 \iff$ i $\leftrightarrow$ j in $G$.
- $U(i,j) = 1 \iff$ i $\rightarrow$ j in $G$.
- $N(i,j) = 1 \iff$ i   j in $G$.

Observation 2: Given 2 nodes $i$, $j$ in $G$, number of $x$ s.t. 

is:

$$\begin{array}{ccc} (i \longleftrightarrow j) & \text{and} & i \rightarrow x \leftarrow j \\ B(i,j) & \times & \sum_{x=1}^{n} U(i,x).U(x,j) \end{array}$$

Royal Academy of Engineering

# Building the Benson Graphs:TO DO

- We are working on an efficient implementation of our formulas.
- Extension to larger motifs ?

# Take Home Messages

- **Partitioning directed networks** : not as "simple" as for undirected case.
- **Motifs** express **complex notions in networks** .
- The **Benson Graph** provides an undirected representation of the network.
- Without other knowledge, $\gamma$-**score** seems to provide good **indication about which motifs used to partition the graph**.
- With linear algebra, the **Benson adjacency matrix** can be **directly built for 3-and 4-node motifs**.

# Thank you for your attention

Some codes and the slides are available on `github.com/luleg/`

**Bibliography**


[1] : *Detection of Terrorism-related Twitter Communities using Centrality Scores*, I.Gialampoukidis et al., 2017
[2] : *Graph Partitioning for Distributed Graph Processing*, M. Onizuka et al, 2017
[3] : *Fast unfolding of communities in large networks*, V. Blondel et al, 2008
[4] : *Multi-Threaded Graph Partitioning*, D. LaSalle, G. Karypis, 2013
[5] : *The map equation*, M. Rosvall et al., 2009
[6] : *Tools for higher-order network analysis*, A. Benson, 2017
[7] : *Higher-order organization of complex networks*, A. Benson et al, 2017
[8] : *Scalable motif-aware graph clustering*, C. Tsourakakis et al, 2017
[9] : *Network Motifs: Simple Building Blocks of Complex Networks*, R. Milo et al, 2002
[10] :*Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities*, A. Lancichinetti, S. Fortunato, 2009
[11] : *gl2vec: Learning Feature Representation Using Graphlets for Directed Networks*, K. Tu et al, 2019
[12] : *FANMOD: a tool for fast network motif detection*, S. Wernicke, F. Rasche, 2006


Royal Academy of Engineering