Ranking and importance in complex networks

Florian Knorn

Florian.Knorn@nuim.ie

Special thanks to Dr. Oliver Mason !

Hamilton Institute

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Institut für Automatisierungstechnik

Otto–von–Guericke Universität

Introduction

Graph theory

Robustness

Real data

Conclusion

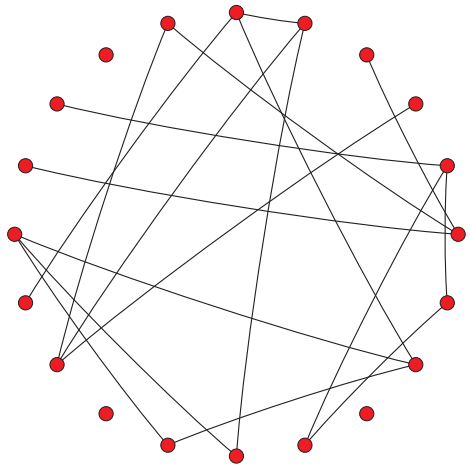
Introduction

Introduction

Introduction
Graph theory
Robustness
Real data
Conclusion

Rough structure of this talk:

- Graph theory *without tears*
  - Basic notions
  - Random graphs
  - Ranking schemes
- Robustness of the ranking schemes
  - Perturbations
  - Deviation measures
  - Results
- Application to real data
  - Datasets
  - Identifying essentiality
  - Results

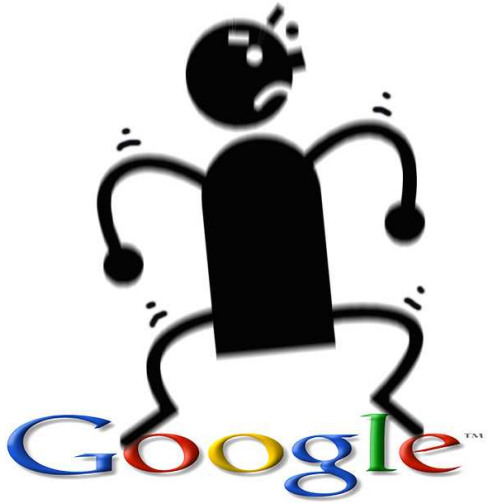


Introduction

Introduction
Graph theory
Robustness
Real data
Conclusion

Rough structure of this talk:

- Graph theory *without tears*
  - Basic notions
  - Random graphs
  - Ranking schemes
- Robustness of the ranking schemes
  - Perturbations
  - Deviation measures
  - Results
- Application to real data
  - Datasets
  - Identifying essentiality
  - Results

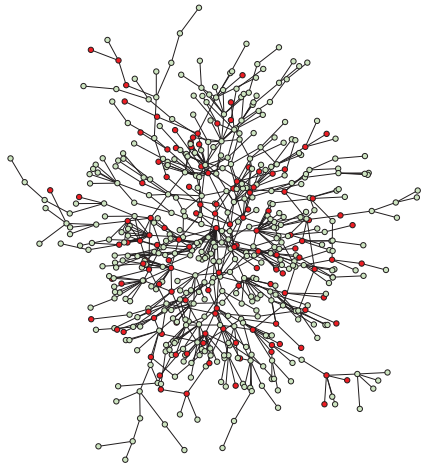


Introduction

Introduction
Graph theory
Robustness
Real data
Conclusion

Rough structure of this talk:

- Graph theory *without tears*
  - Basic notions
  - Random graphs
  - Ranking schemes
- Robustness of the ranking schemes
  - Perturbations
  - Deviation measures
  - Results
- Application to real data
  - Datasets
  - Identifying essentiality
  - Results



Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Graph theory

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Introduction

- Idea by Euler, $\sim$ 1730

  $\vdots$

- Some of the applications today: modeling of
  - Internet, WWW, Intranets
  - Social networks, epidemics
  - Infrastructures (roads, power grids)
  - Phone call networks
  - Collaboration networks
  - Protein–protein interaction networks

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Introduction

- Idea by Euler, $\sim$ 1730

  ⋮

- Some of the applications today: modeling of
  - Internet, WWW, Intranets
  - Social networks, epidemics
  - Infrastructures (roads, power grids)
  - Phone call networks
  - Collaboration networks
  - Protein–protein interaction networks

Basic notions

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Basic notions

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

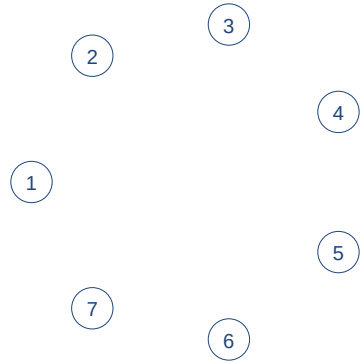
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

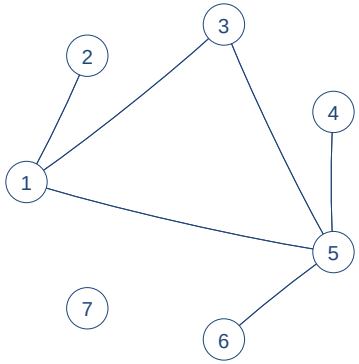
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E})$ ...
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

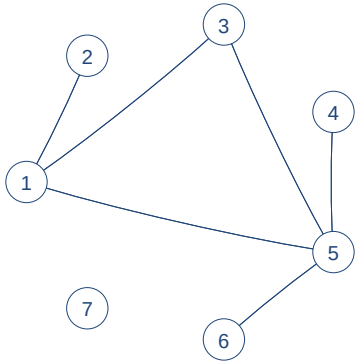
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

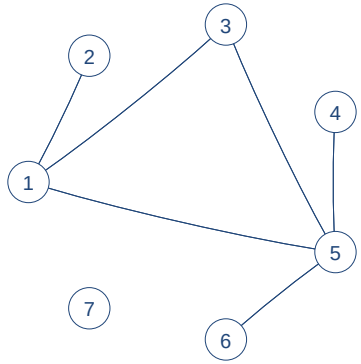
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

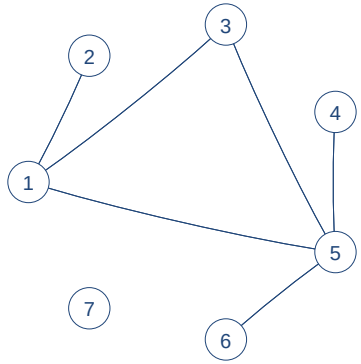
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

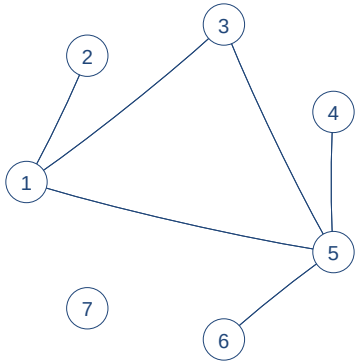
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

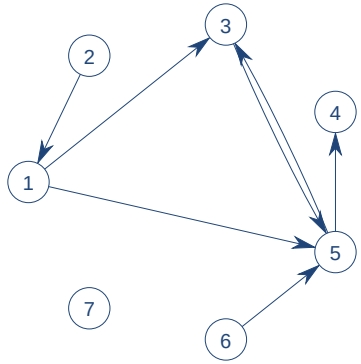
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

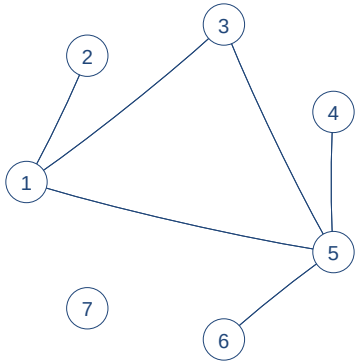
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

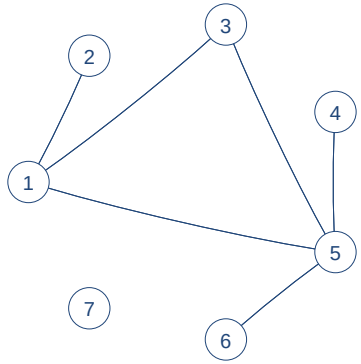
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

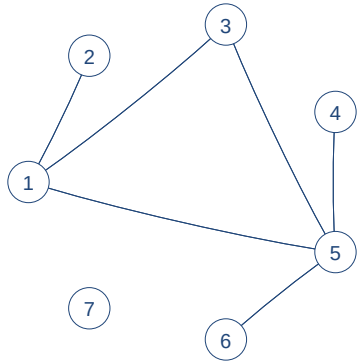
Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Basic notions



- What is a graph?
  - Loosely speaking: *A graph is a bunch of nodes connected by a bunch of edges*
  - More formally: $\mathscr{G} = (\mathcal{V}, \mathcal{E}) \ldots$
  - Description via matrices: adjacency–, distance–, Laplacian– & incidence–matrix
- Characteristics
  - Undirected / directed
  - Connected / disconnected
  - Characteristic values: number of nodes / edges, average node degree, degree distribution, average path length, diameter, clustering coefficient, ...

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

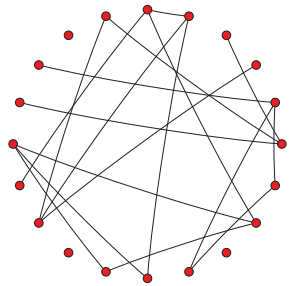
Real data

Conclusion

Random graphs

We looked at 3 models:

- Erdös–Rényi (classical), $\sim$ 1957
  - Seminal, first model
  - No "order"
  - Few practical applications



- Watts–Strogatz (small–world) $\sim$ 1999
  - Model for social networks
  - Highly clustered, rel. short paths
  - Used in sociology, biology, ...
- Barabási–Albert (scale–free) $\sim$ 1999
  - Model for networks with growth
  - Very short paths, few hub nodes
  - Many practical applications

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

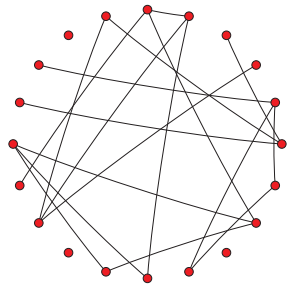
Real data

Conclusion

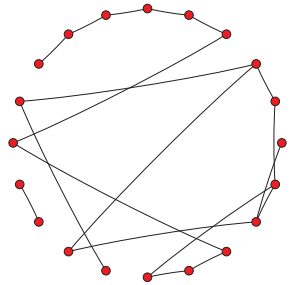
Random graphs

We looked at 3 models:

- Erdös–Rényi (classical), $\sim$ 1957
  - Seminal, first model
  - No "order"
  - Few practical applications



- Watts–Strogatz (small–world) $\sim$ 1999
  - Model for social networks
  - Highly clustered, rel. short paths
  - Used in sociology, biology, ...



- Barabási–Albert (scale–free) $\sim$ 1999
  - Model for networks with growth
  - Very short paths, few hub nodes
  - Many practical applications

Random graphs

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

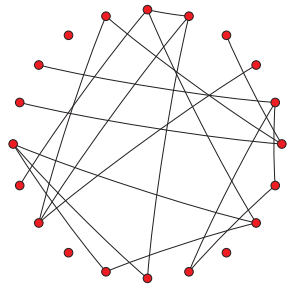
Robustness

Real data

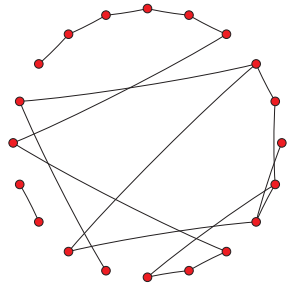
Conclusion

We looked at 3 models:

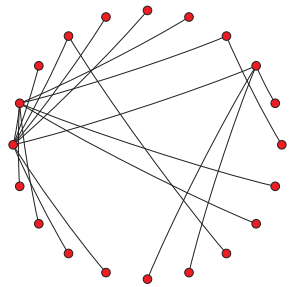
- Erdös–Rényi (classical), $\sim$ 1957
  - Seminal, first model
  - No "order"
  - Few practical applications



- Watts–Strogatz (small–world) $\sim$ 1999
  - Model for social networks
  - Highly clustered, rel. short paths
  - Used in sociology, biology, ...



- Barabási–Albert (scale–free) $\sim$ 1999
  - Model for networks with growth
  - Very short paths, few hub nodes
  - Many practical applications



Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Ranking schemes

- Growing complexity of networks — *Who's important?*

  ⇒ Ranking of the nodes
- Most immediate
  - Node degrees (ND)
- Eigenvector based
  - HITS
  - PageRank (PR)
- Centrality based
  - Excentricity (EXC)
  - Status (STA)
  - Centroid value (CV)
- Impact on topology based
  - Damage (DAM)

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Ranking schemes

- Growing complexity of networks — *Who's important?*

  ⇒ Ranking of the nodes
- Most immediate
  - Node degrees (ND)
- Eigenvector based
  - HITS
  - PageRank (PR)
- Centrality based
  - Excentricity (EXC)
  - Status (STA)
  - Centroid value (CV)
- Impact on topology based
  - Damage (DAM)

Place	Score	Node
1.	0.2568	G
2.	0.1644	A
3.	0.1118	C
4.	0.1053	B
5.	0.0625	E
6.	0.0616	F
7.	0.0616	D
⋮	⋮	⋮

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Ranking schemes

- Growing complexity of networks — *Who's important?*

  ⇒ Ranking of the nodes
- Most immediate
  - Node degrees (ND)
- Eigenvector based
  - HITS
  - PageRank (PR)
- Centrality based
  - Excentricity (EXC)
  - Status (STA)
  - Centroid value (CV)
- Impact on topology based
  - Damage (DAM)

Place	Score	Node
1.	0.2568	G
2.	0.1644	A
3.	0.1118	C
4.	0.1053	B
5.	0.0625	E
6.	0.0616	F
7.	0.0616	D
⋮	⋮	⋮

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Ranking schemes

- Growing complexity of networks — *Who's important?*

  ⇒ Ranking of the nodes
- Most immediate
  - Node degrees (ND)
- Eigenvector based
  - HITS
  - PageRank (PR)
- Centrality based
  - Excentricity (EXC)
  - Status (STA)
  - Centroid value (CV)
- Impact on topology based
  - Damage (DAM)

Place	Score	Node
1.	0.2568	G
2.	0.1644	A
3.	0.1118	C
4.	0.1053	B
5.	0.0625	E
6.	0.0616	F
7.	0.0616	D
⋮	⋮	⋮

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Ranking schemes

- Growing complexity of networks — *Who's important?*

  ⇒ Ranking of the nodes
- Most immediate
  - Node degrees (ND)
- Eigenvector based
  - HITS
  - PageRank (PR)
- Centrality based
  - Excentricity (EXC)
  - Status (STA)
  - Centroid value (CV)
- Impact on topology based
  - Damage (DAM)

Place	Score	Node
1.	0.2568	G
2.	0.1644	A
3.	0.1118	C
4.	0.1053	B
5.	0.0625	E
6.	0.0616	F
7.	0.0616	D
⋮	⋮	⋮

Introduction

Graph theory
· Introduction
· Basic notions
· Rand. graphs
· Ranking schemes

Robustness

Real data

Conclusion

Ranking schemes

- Growing complexity of networks — *Who's important?*

  ⇒ Ranking of the nodes
- Most immediate
  - Node degrees (ND)
- Eigenvector based
  - HITS
  - PageRank (PR)
- Centrality based
  - Excentricity (EXC)
  - Status (STA)
  - Centroid value (CV)
- Impact on topology based
  - Damage (DAM)

Place	Score	Node
1.	0.2568	G
2.	0.1644	A
3.	0.1118	C
4.	0.1053	B
5.	0.0625	E
6.	0.0616	F
7.	0.0616	D
⋮	⋮	⋮

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Robustness

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	C
2.	A
3.	B

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	B
2.	A ?
3.	C

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	B
2.	A ?
3.	C

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	B
2.	A ?
3.	C

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	B
2.	A ?
3.	C

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	B
2.	A ?
3.	C

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	B
2.	A ?
3.	C

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	B
2.	A ?
3.	C

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	B
2.	A ?
3.	C

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Introduction

- What impact have perturbations on rankings?
- Real data in many cases noisy / missing:
  - Inaccurate measuring
  - Deliberate sampling
- 6 deviation measures:
  - 3 based on node ranks
  - 3 based on top percentile of nodes
- Types of perturbation:
  - Edge removal / rewiring / addition
  - Node removal
- Simulations:
  1. Generate & rank BA graph
  2. Perturb graph
  3. Rank perturbed graph
  4. Compare rankings

Place	Node
1.	B
2.	A ?
3.	C

Deviation measures

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

- Based on node ranks:

Original		Perturbed	
Rank	Node	Rank	Node
1.	C		
2.	H		
3.	D		
4.	J		
5.	A		
.	.		
.	.		
.	.		
.	.		
.	.		

- Based on top fraction of nodes:

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Deviation measures

- Based on node ranks:

Original		Perturbed	
Rank	Node	Rank	Node
1.	C	1	A
2.	H	2	H
3.	D	.	.
4.	J	4	J
5.	A	.	.
.	.	6	C
.	.	.	.
.	.	.	.
.	.	9	D
.	.	.	.

- Based on top fraction of nodes:

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Deviation measures

- Based on node ranks:

Original		Perturbed	
Rank	Node	Rank	Node
1.	C	1	A
2.	H	2	H
3.	D	.	.
4.	J	4	J
5.	A	.	.
.	.	6	C
.	.	.	.
.	.	.	.
.	.	9	D
.	.	.	.

$$\delta_1 = \underbrace{|1-6|}_{\text{Node C}} + \underbrace{|2-2|}_{\text{Node H}} + \underbrace{|3-9|}_{\text{Node D}} + \underbrace{|4-4|}_{\text{Node J}} + \underbrace{|5-1|}_{\text{Node A}} = 15$$

- Based on top fraction of nodes:

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Deviation measures

- Based on node ranks:

Original		Perturbed	
Rank	Node	Rank	Node
1.	C	1	A
2.	H	2	H
3.	D	.	.
4.	J	4	J
5.	A	.	.
.	.	6	C
.	.	.	.
.	.	.	.
.	.	9	D
.	.	.	.

$$\delta_1 = \underbrace{|1-6|}_{\text{Node C}} + \underbrace{|2-2|}_{\text{Node H}} + \underbrace{|3-9|}_{\text{Node D}} + \underbrace{|4-4|}_{\text{Node J}} + \underbrace{|5-1|}_{\text{Node A}} = 15$$

- Based on top fraction of nodes:

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

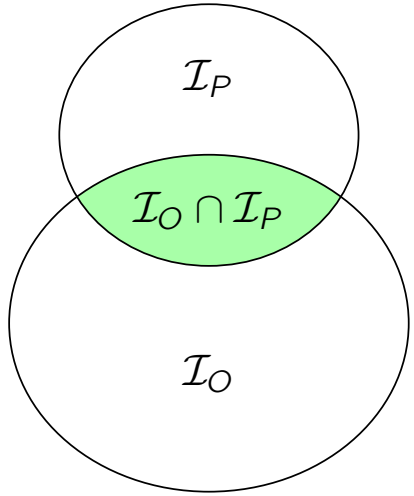
Deviation measures

- Based on node ranks:

Original		Perturbed	
Rank	Node	Rank	Node
1.	C	1	A
2.	H	2	H
3.	D	.	.
4.	J	4	J
5.	A	.	.
.	.	6	C
.	.	.	.
.	.	.	.
.	.	9	D
.	.	.	.

$$\delta_1 = \underbrace{|1-6|}_{\text{Node C}} + \underbrace{|2-2|}_{\text{Node H}} + \underbrace{|3-9|}_{\text{Node D}} + \underbrace{|4-4|}_{\text{Node J}} + \underbrace{|5-1|}_{\text{Node A}} = 15$$

- Based on top fraction of nodes:



Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Deviation measures

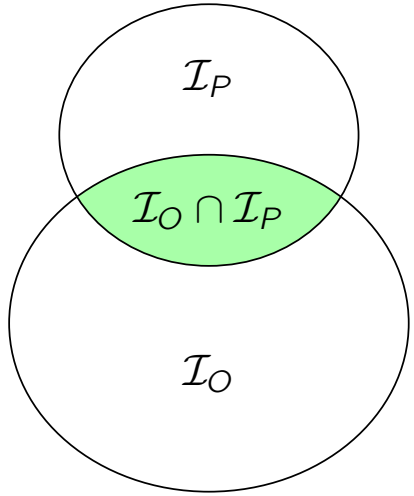
- Based on node ranks:

Original		Perturbed	
Rank	Node	Rank	Node
1.	C	1	A
2.	H	2	H
3.	D	.	.
4.	J	4	J
5.	A	.	.
.	.	6	C
.	.	.	.
.	.	.	.
.	.	9	D
.	.	.	.

$$\delta_1 = \underbrace{|1-6|}_{\text{Node C}} + \underbrace{|2-2|}_{\text{Node H}} + \underbrace{|3-9|}_{\text{Node D}} + \underbrace{|4-4|}_{\text{Node J}} + \underbrace{|5-1|}_{\text{Node A}} = 15$$

- Based on top fraction of nodes:

$$\delta_4 = 1 - \frac{|\mathcal{I}_O \cap \mathcal{I}_P|}{|\mathcal{I}_P|}$$



Introduction

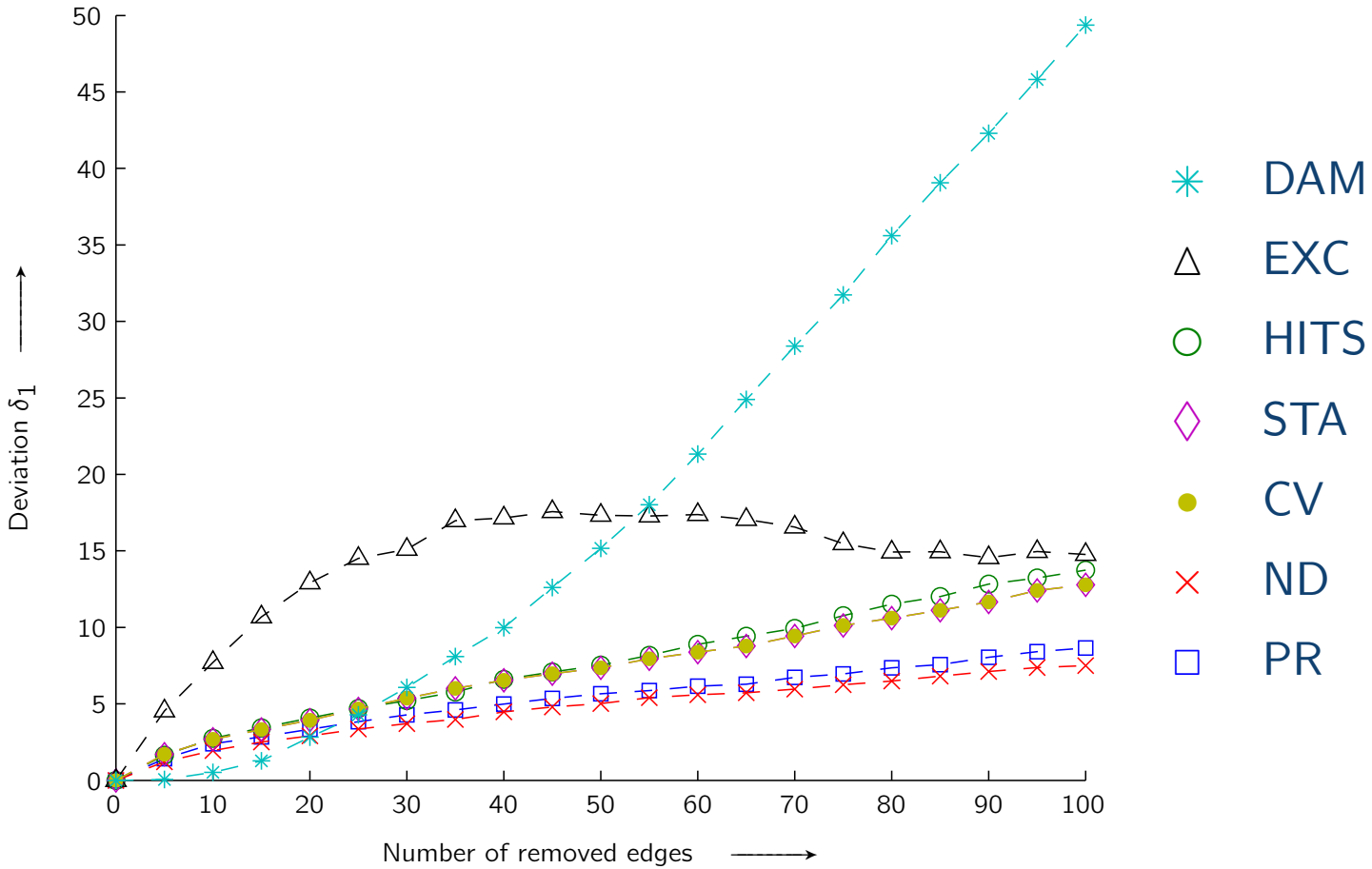
Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

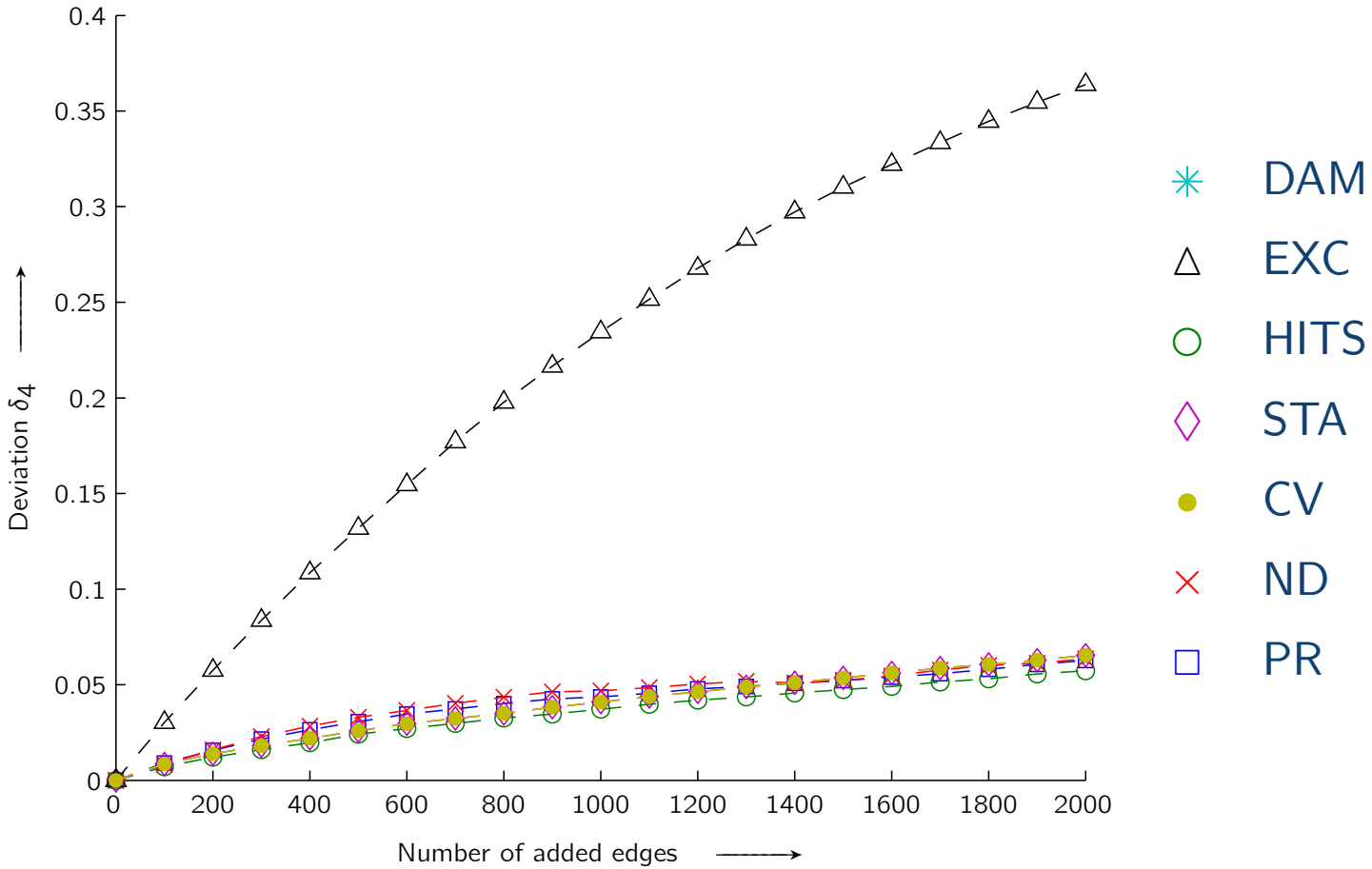
Results



Average value of 250 Barabási–Albert graphs on 150 nodes and 450 edges

Introduction
Graph theory
Robustness
· Introduction
· Dev. measures
· Results
Real data
Conclusion

Results



Average value of 250 Barabási–Albert graphs on 3000 nodes and 9000 edges

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Results

- PR and ND often most robust
- However, few cases where HITS, CV or STA better
- EXC sensitive + useless when graph disconnected
- DAM relatively robust, but costly
- Choice always depends on:
  - Interpretation of "deviation"
  - Type of perturbations
  - Resolution of the ranking needed
  - Interpretation of "importance" $\Rightarrow$ application

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Results

- PR and ND often most robust
- However, few cases where HITS, CV or STA better
- EXC sensitive + useless when graph disconnected
- DAM relatively robust, but costly
- Choice always depends on:
  - Interpretation of "deviation"
  - Type of perturbations
  - Resolution of the ranking needed
  - Interpretation of "importance" $\Rightarrow$ application

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Results

- PR and ND often most robust
- However, few cases where HITS, CV or STA better
- EXC sensitive + useless when graph disconnected
- DAM relatively robust, but costly
- Choice always depends on:
  - Interpretation of "deviation"
  - Type of perturbations
  - Resolution of the ranking needed
  - Interpretation of "importance" $\Rightarrow$ application

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Results

- PR and ND often most robust
- However, few cases where HITS, CV or STA better
- EXC sensitive + useless when graph disconnected
- DAM relatively robust, but costly
- Choice always depends on:
  - Interpretation of "deviation"
  - Type of perturbations
  - Resolution of the ranking needed
  - Interpretation of "importance" $\Rightarrow$ application

Introduction

Graph theory

Robustness
· Introduction
· Dev. measures
· Results

Real data

Conclusion

Results

- PR and ND often most robust
- However, few cases where HITS, CV or STA better
- EXC sensitive + useless when graph disconnected
- DAM relatively robust, but costly
- Choice always depends on:
  - Interpretation of "deviation"
  - Type of perturbations
  - Resolution of the ranking needed
  - Interpretation of "importance" $\Rightarrow$ application

Introduction

Graph theory

Robustness

Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness

Conclusion

Real data

Introduction

Graph theory

Robustness

Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness

Conclusion

Introduction

- Nodes' ranked importance $\overset{?}{\leftrightsquigarrow}$ real importance ?

- Protein–Protein interaction networks
  - Undirected graphs
  - Isolate largest connected component
  - Top–ranked–nodes $\overset{?}{\leftrightsquigarrow}$ Essential–for–survival

- Measure of success:

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Introduction

- Nodes' ranked importance $\overset{?}{\leftrightsquigarrow}$ real importance ?
  $\Rightarrow$ Depends on the context, the *application*

- Protein–Protein interaction networks
  - Undirected graphs
  - Isolate largest connected component
  - Top–ranked–nodes $\overset{?}{\leftrightsquigarrow}$ Essential–for–survival

- Measure of success:

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Introduction

- Nodes' ranked importance $\overset{?}{\leftrightsquigarrow}$ real importance ?
  $\Rightarrow$ Depends on the context, the *application*

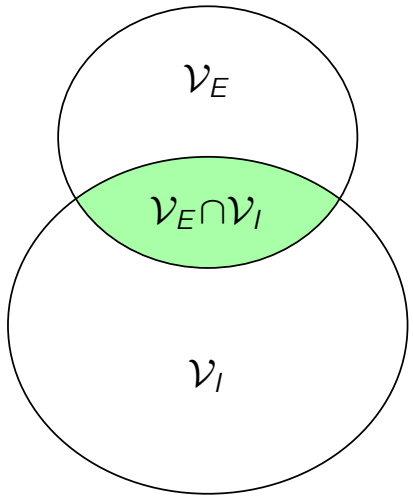
- Protein–Protein interaction networks
  - Undirected graphs
  - Isolate largest connected component
  - Top–ranked–nodes $\overset{?}{\leftrightsquigarrow}$ Essential–for–survival

- Measure of success:

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Introduction

- Nodes' ranked importance $\overset{?}{\leftrightsquigarrow}$ real importance ?
  $\Rightarrow$ Depends on the context, the *application*
- Protein–Protein interaction networks
  - Undirected graphs
  - Isolate largest connected component
  - Top–ranked–nodes $\overset{?}{\leftrightsquigarrow}$ Essential–for–survival
- Measure of success:



Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

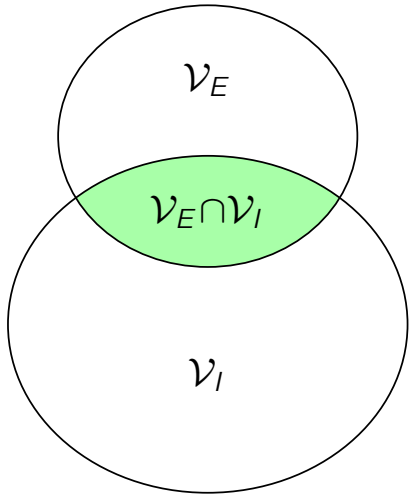
Introduction

- Nodes' ranked importance $\overset{?}{\leftrightsquigarrow}$ real importance ?
  $\Rightarrow$ Depends on the context, the *application*

- Protein–Protein interaction networks
  - Undirected graphs
  - Isolate largest connected component
  - Top–ranked–nodes $\overset{?}{\leftrightsquigarrow}$ Essential–for–survival

- Measure of success:

$$h = \frac{|\mathcal{V}_E \cap \mathcal{V}_I|}{|\mathcal{V}_I|}$$



Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

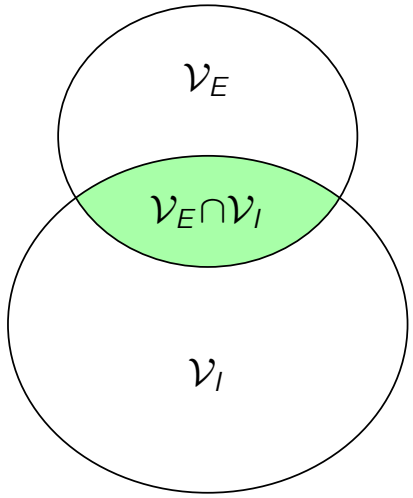
Introduction

- Nodes' ranked importance $\overset{?}{\leftrightsquigarrow}$ real importance ?
  $\Rightarrow$ Depends on the context, the *application*

- Protein–Protein interaction networks
  - Undirected graphs
  - Isolate largest connected component
  - Top–ranked–nodes $\overset{?}{\leftrightsquigarrow}$ Essential–for–survival

- Measure of success:

$$h = \frac{|\mathcal{V}_E \cap \mathcal{V}_I|}{|\mathcal{V}_I|} \overset{!}{>} h_{ess}$$



Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Datasets

Protein–protein interaction network of the budding yeast *Saccharomyces cerevisiae*.

- Uetz *et al.* [1)]
  - $n = 558, \quad e = 646$
  - $h_{ess} = 22.6\%$
- Ito *et al.* [2)]
  - $n = 2840, \quad e = 4147$
  - $h_{ess} = 17.9\%$
- Yu *et al.* [3)]
  - $n = 4544, \quad e = 22587$
  - $h_{ess} = 21.0\%$

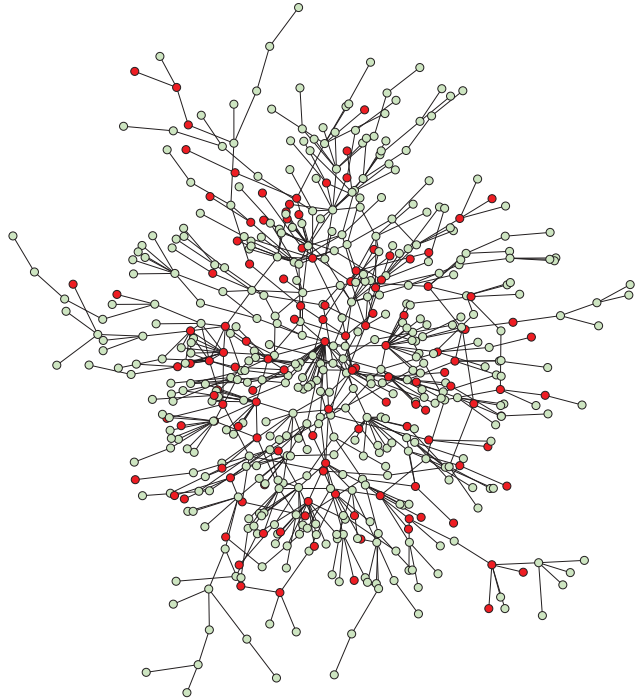
[1)] *Nature*, 403:623–627, Feb. 2000
[2)] *Proc. of the Nat. Academy of Sciences of the USA*, 98:4569–4574, April 2001
[3)] *Trends in Genetics*, 20(6):227–231, June 2004

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Datasets

Protein–protein interaction network of the budding yeast *Saccharomyces cerevisiae*.

- Uetz *et al.* [1)]
  - $n = 558, \quad e = 646$
  - $h_{ess} = 22.6\%$
- Ito *et al.* [2)]
  - $n = 2840, \quad e = 4147$
  - $h_{ess} = 17.9\%$
- Yu *et al.* [3)]
  - $n = 4544, \quad e = 22587$
  - $h_{ess} = 21.0\%$



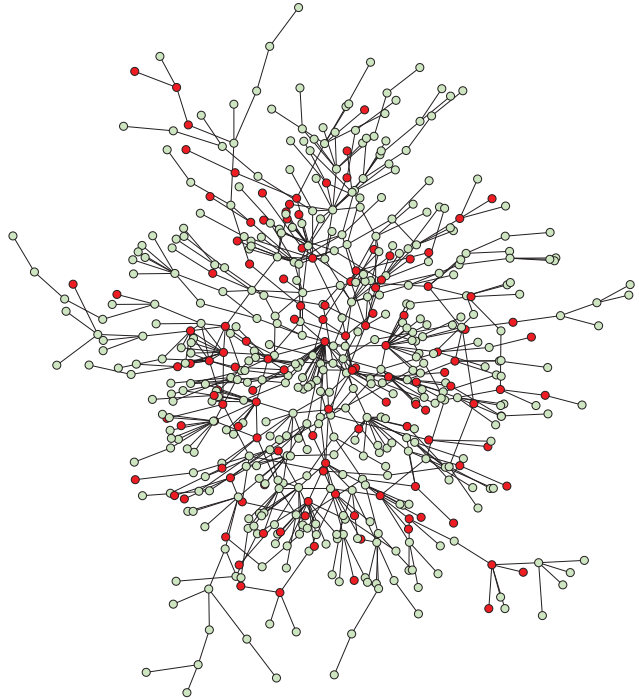
[1)] *Nature*, 403:623–627, Feb. 2000
[2)] *Proc. of the Nat. Academy of Sciences of the USA*, 98:4569–4574, April 2001
[3)] *Trends in Genetics*, 20(6):227–231, June 2004

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Datasets

Protein–protein interaction network of the budding yeast *Saccharomyces cerevisiae*.

- Uetz *et al.* [1)]
  - $n = 558, \quad e = 646$
  - $h_{ess} = 22.6\%$
- Ito *et al.* [2)]
  - $n = 2840, \quad e = 4147$
  - $h_{ess} = 17.9\%$
- Yu *et al.* [3)]
  - $n = 4544, \quad e = 22587$
  - $h_{ess} = 21.0\%$



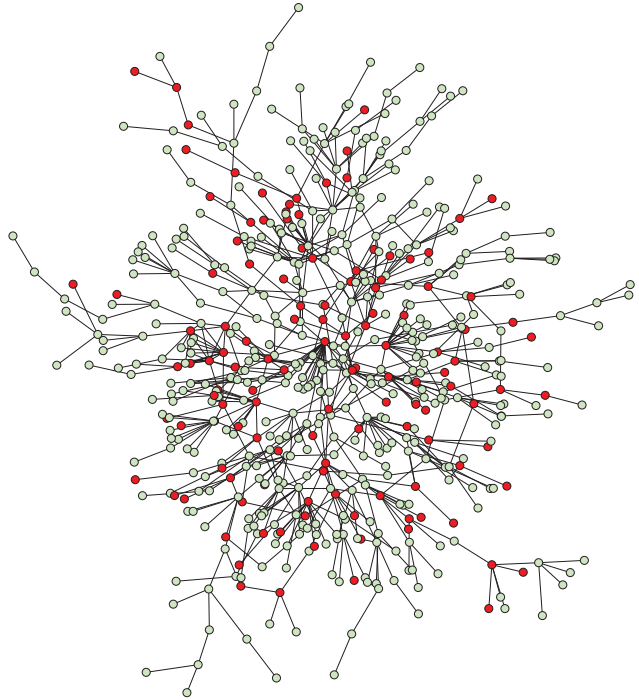
[1)] *Nature*, 403:623–627, Feb. 2000
[2)] *Proc. of the Nat. Academy of Sciences of the USA*, 98:4569–4574, April 2001
[3)] *Trends in Genetics*, 20(6):227–231, June 2004

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Datasets

Protein–protein interaction network of the budding yeast *Saccharomyces cerevisiae*.

- Uetz *et al.* [1)]
  - $n = 558, \quad e = 646$
  - $h_{ess} = 22.6\%$
- Ito *et al.* [2)]
  - $n = 2840, \quad e = 4147$
  - $h_{ess} = 17.9\%$
- Yu *et al.* [3)]
  - $n = 4544, \quad e = 22587$
  - $h_{ess} = 21.0\%$



[1)] *Nature*, 403:623–627, Feb. 2000
[2)] *Proc. of the Nat. Academy of Sciences of the USA*, 98:4569–4574, April 2001
[3)] *Trends in Genetics*, 20(6):227–231, June 2004

Identifying essentiality

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

- We considered the top ranked 5% of nodes as "important"
- PR & ND most adept
- STA & CV mixed results
- Further investigations needed for DAM

Rank. sch.	Uetz	Ito	Yu
ND	**48.4**	23.8	42.3
HITS	35.7	16.6	43.8
PR	46.4	**24.8**	**62.1**
EXC	24.6	18.4	34.7
STA	17.2	19.7	59.5
CV	17.2	19.7	59.5
DAM	44.0	?	?
h_{ess}	22.6	17.9	21.0

Identifying essentiality

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

- We considered the top ranked 5% of nodes as "important"
- PR & ND most adept
- STA & CV mixed results
- Further investigations needed for DAM

Rank. sch.	Uetz	Ito	Yu
ND	**48.4**	23.8	42.3
HITS	35.7	16.6	43.8
PR	46.4	**24.8**	**62.1**
EXC	24.6	18.4	34.7
STA	17.2	19.7	59.5
CV	17.2	19.7	59.5
DAM	44.0	?	?
h_{ess}	22.6	17.9	21.0

Identifying essentiality

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

- We considered the top ranked 5% of nodes as "important"
- PR & ND most adept
- STA & CV mixed results
- Further investigations needed for DAM

Rank. sch.	Uetz	Ito	Yu
ND	**48.4**	23.8	42.3
HITS	35.7	16.6	43.8
PR	46.4	**24.8**	**62.1**
EXC	24.6	18.4	34.7
STA	17.2	19.7	59.5
CV	17.2	19.7	59.5
DAM	44.0	?	?
h_{ess}	22.6	17.9	21.0

Identifying essentiality

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

- We considered the top ranked 5% of nodes as "important"
- PR & ND most adept
- STA & CV mixed results
- Further investigations needed for DAM

Rank. sch.	Uetz	Ito	Yu
ND	**48.4**	23.8	42.3
HITS	35.7	16.6	43.8
PR	46.4	**24.8**	**62.1**
EXC	24.6	18.4	34.7
STA	17.2	19.7	59.5
CV	17.2	19.7	59.5
DAM	44.0	?	?
h_{ess}	22.6	17.9	21.0

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Robustness

- Data surely inaccurate to some extent
- Study the effect of increasing (edge) perturbations
- PR and ND still best results
- But HITS most robust (= least relative change)

Rank. sch.	Uetz	Ito	Yu
ND	46.3	23.0	43.3
HITS	39.7	*17.0*	44.5
PR	**47.0**	**23.8**	**64.1**
EXC	*19.6*	18.2	29.2
STA	*19.4*	20.3	57.1
CV	*19.4*	20.3	57.1
DAM	36.2	?	?
h_{ess}	22.6	17.9	21.0

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Robustness

- Data surely inaccurate to some extent
- Study the effect of increasing (edge) perturbations
- PR and ND still best results
- But HITS most robust (= least relative change)

Rank. sch.	Uetz	Ito	Yu
ND	46.3	23.0	43.3
HITS	39.7	*17.0*	44.5
PR	**47.0**	**23.8**	**64.1**
EXC	*19.6*	18.2	29.2
STA	*19.4*	20.3	57.1
CV	*19.4*	20.3	57.1
DAM	36.2	?	?
h_{ess}	22.6	17.9	21.0

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Robustness

- Data surely inaccurate to some extent
- Study the effect of increasing (edge) perturbations
- PR and ND still best results
- But HITS most robust (= least relative change)

Rank. sch.	Uetz	Ito	Yu
ND	46.3	23.0	43.3
HITS	39.7	*17.0*	44.5
PR	**47.0**	**23.8**	**64.1**
EXC	*19.6*	18.2	29.2
STA	*19.4*	20.3	57.1
CV	*19.4*	20.3	57.1
DAM	36.2	?	?
h_{ess}	22.6	17.9	21.0

Introduction
Graph theory
Robustness
Real data
· Introduction
· Datasets
· Identif. essent.
· Robustness
Conclusion

Robustness

- Data surely inaccurate to some extent
- Study the effect of increasing (edge) perturbations
- PR and ND still best results
- But HITS most robust (= least relative change)

Rank. sch.	Uetz	Ito	Yu
ND	46.3	23.0	43.3
HITS	39.7	*17.0*	44.5
PR	**47.0**	**23.8**	**64.1**
EXC	*19.6*	18.2	29.2
STA	*19.4*	20.3	57.1
CV	*19.4*	20.3	57.1
DAM	36.2	?	?
h_{ess}	22.6	17.9	21.0

Introduction
Graph theory
Robustness
Real data
Conclusion

Conclusion

Introduction
Graph theory
Robustness
Real data
Conclusion

Conclusion

- Developed large collection of MATLAB functions for graph theory, random graphs & ranking schemes
- Robustness of the ranking schemes
  - PR and ND *usually* most robust
  - EXC least suitable
- Application to real data:
  - PR and ND again *usually* most suitable
  - HITS most robust
- ND apparently very good approx. to more costly PR
- Recommended further investigations:
  - Other ranking schemes, more investig. for damage
  - Use different deviation measures
  - More structured perturbations
  - Other datasets, directed graphs

Conclusion

Introduction
Graph theory
Robustness
Real data
Conclusion

- Developed large collection of MATLAB functions for graph theory, random graphs & ranking schemes
- Robustness of the ranking schemes
  - PR and ND *usually* most robust
  - EXC least suitable
- Application to real data:
  - PR and ND again *usually* most suitable
  - HITS most robust
- ND apparently very good approx. to more costly PR
- Recommended further investigations:
  - Other ranking schemes, more investig. for damage
  - Use different deviation measures
  - More structured perturbations
  - Other datasets, directed graphs

Introduction
Graph theory
Robustness
Real data
Conclusion

Conclusion

- Developed large collection of MATLAB functions for graph theory, random graphs & ranking schemes
- Robustness of the ranking schemes
  - PR and ND *usually* most robust
  - EXC least suitable
- Application to real data:
  - PR and ND again *usually* most suitable
  - HITS most robust
- ND apparently very good approx. to more costly PR
- Recommended further investigations:
  - Other ranking schemes, more investig. for damage
  - Use different deviation measures
  - More structured perturbations
  - Other datasets, directed graphs

Conclusion

Introduction
Graph theory
Robustness
Real data
Conclusion

- Developed large collection of MATLAB functions for graph theory, random graphs & ranking schemes
- Robustness of the ranking schemes
  - PR and ND *usually* most robust
  - EXC least suitable
- Application to real data:
  - PR and ND again *usually* most suitable
  - HITS most robust
- ND apparently very good approx. to more costly PR
- Recommended further investigations:
  - Other ranking schemes, more investig. for damage
  - Use different deviation measures
  - More structured perturbations
  - Other datasets, directed graphs

Introduction
Graph theory
Robustness
Real data
Conclusion

Conclusion

- Developed large collection of MATLAB functions for graph theory, random graphs & ranking schemes
- Robustness of the ranking schemes
  - PR and ND *usually* most robust
  - EXC least suitable
- Application to real data:
  - PR and ND again *usually* most suitable
  - HITS most robust
- ND apparently very good approx. to more costly PR
- Recommended further investigations:
  - Other ranking schemes, more investig. for damage
  - Use different deviation measures
  - More structured perturbations
  - Other datasets, directed graphs

Many thanks to

- Dr. Oliver Mason, Prof. Robert Shorten & all "Hamiltonians"
- Prof. Jörg Raisch & Janine Holzmann
- Studienstiftung & Science Foundation Ireland (grant no. 04/In1/I478)
- My family and close friends

Hamilton Institute

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Institut für Automatisierungstechnik

Otto–von–Guericke Universität

Many thanks to

- Dr. Oliver Mason, Prof. Robert Shorten & all "Hamiltonians"
- Prof. Jörg Raisch & Janine Holzmann
- Studienstiftung & Science Foundation Ireland (grant no. 04/In1/I478)
- My family and close friends

Hamilton Institute

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Institut für Automatisierungstechnik

Otto–von–Guericke Universität

Many thanks to

- Dr. Oliver Mason, Prof. Robert Shorten & all "Hamiltonians"
- Prof. Jörg Raisch & Janine Holzmann
- Studienstiftung & Science Foundation Ireland (grant no. 04/In1/I478)
- My family and close friends

Hamilton Institute

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Institut für Automatisierungstechnik

Otto–von–Guericke Universität

Many thanks to

- Dr. Oliver Mason, Prof. Robert Shorten & all "Hamiltonians"
- Prof. Jörg Raisch & Janine Holzmann
- Studienstiftung & Science Foundation Ireland (grant no. 04/In1/I478)
- My family and close friends

Hamilton Institute

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Institut für Automatisierungstechnik

Otto–von–Guericke Universität

Many thanks to

- Dr. Oliver Mason, Prof. Robert Shorten & all "Hamiltonians"
- Prof. Jörg Raisch & Janine Holzmann
- Studienstiftung & Science Foundation Ireland (grant no. 04/In1/I478)
- My family and close friends
- You for being such a nice and considering audience :)

Hamilton Institute

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Institut für Automatisierungstechnik

Otto–von–Guericke Universität