Analysis of Complex Networks: Applications and Challenges

M. Dehmer

Berner Fachhochschule
UMIT, The Health and Life Sciences University, Austria

Outline

Part I

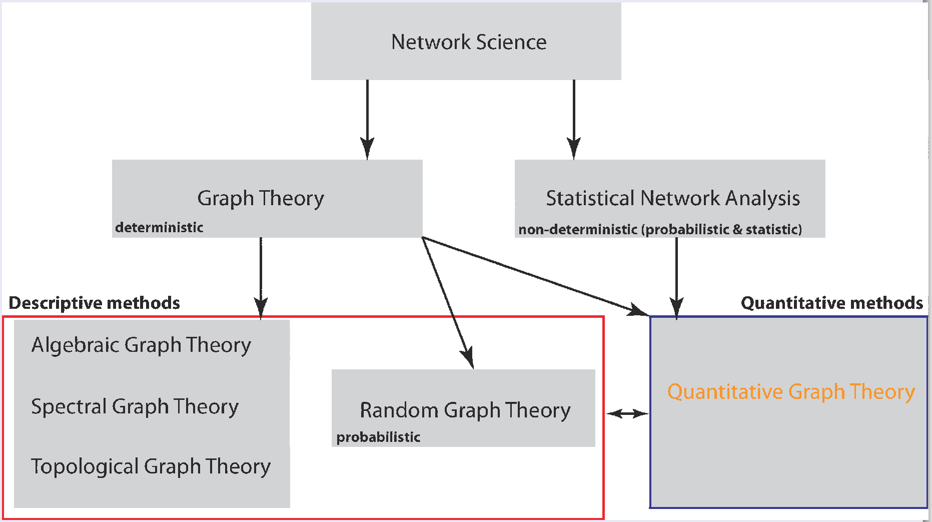
Brief Introduction

Application of Data Science: Quantitative Network Analysis

Various graph-based techniques have been developed. For example:

- Graph classes such as small-world and scale free to characterize real-world networks, e.g., WWW etc. (Newman, 2012)
- Graph Mining-techniques such as frequent patterns, motif search, shortest path analysis, and so forth
- Graph measures based on distances, vertex degrees, eigenvalues and entropy (see, e.g., Dehmer, Chen, Shi, 2020)
- Classical graph measures often possess inefficient time complexity
- An important problem of structural data analysis is to generate the networks exhaustively

Overview about Network Science



see (Dehmer, Emmert-Streib, 2018)

Application I: Analysis of Transportation Networks

- A transportation network is a graph $G = (V, E)$ where V are the vertices (e.g., stations, airports etc.) and E connections between those vertices (train or flight connections etc.)
    - What kind of structural features of a transportation network give risk factors?
    - To quantify structural information, one needs a quantitative approach
    - A quantitative network measure is a mapping $I : \mathcal{G} \longrightarrow \mathbb{R}_+$
    - Prominent examples are the Wiener index or degree measures given by $W(G) := \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d(v_i, v_j)$ or $D(v_i) := \delta(v_i)$
    - Which measure is the most efficient one?
    - Problem: How vulnerable are transportation networks?
    - Efficient approaches are needed to estimate the possibility of threat (Big Data!)

Software-based Approach

Problem: Finding efficient vulnerability measures, see Dehmer et al. (2013, 2018)

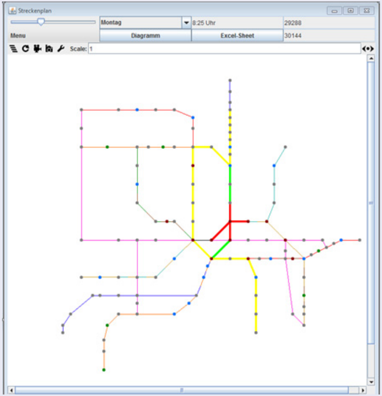


Figure: The Munich subway network and possible risk factors

Application II: Stock Market Data Analysis

Goal: To avoid getting broke after the Lehman Brothers-disaster

- Most of the contributions deal with analyzing stocks one by one (one dimensional)
- Emmert-Streib and Dehmer (2014) found that relationships between stocks are crucial
- They inferred financial networks from complete NASDAQ-data for a long time interval
- They calculated a so-called reference graph G^r and defined comparative graph measure

$$d(t) = d(G^t, G^r) \quad \forall\, t.$$

- The interpretation of G^t_{ij} is the probability that stock i and stock j are correlated in the considered time intervals

Result: Financial Crash Detection

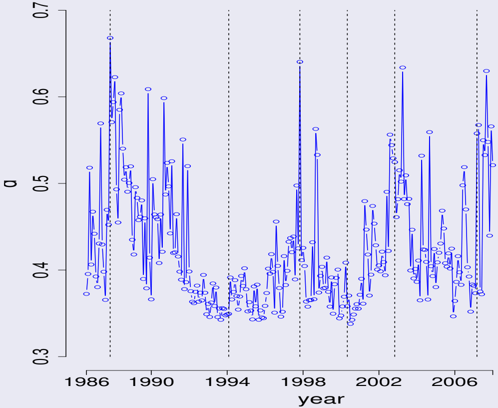


Figure: Exploratory Data Analysis: Financial Crash Detection, Emmert-Streib and Dehmer (2014)

Structural Network Descriptors – Introduction

How can we quantify the structure of a network?

- Remind that a topological descriptor (measure) is a mapping $I : \mathcal{G} \longrightarrow \mathbb{R}_+$
- Several groups of descriptors exist, e.g., information-theoretic, non-information-theoretic, distance-based etc.
- In particular, properties of information-theoretic measures have been explored extensively:
  - Chen Z., Dehmer M., Shi Y.: Bounds for degree-based Network Entropies, Applied Mathematics and Computation, Vol. 265, 2015, 983-993
  - Chen Z., Dehmer M., Emmert-Streib F., Shi Y.: Entropy of Weighted Graphs with Randic Weights, Entropy, Vol. 17 (6), 2015, 3710-3723

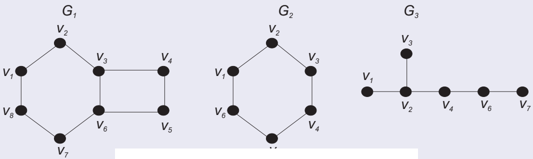
Structural Network Descriptors – Wiener and Randić Index

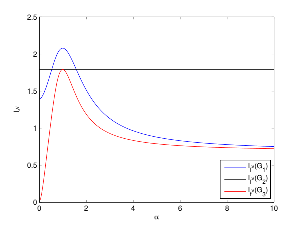
- Prominent examples are the Wiener index and Randić index given by $W(G) := \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d(v_i, v_j)$ and $R(G) := \sum_{(v_i,v_j)\in E}[k_{v_i}k_{v_j}]^{-\frac{1}{2}}$
- W and R have extensively been used to predict physico-chemical properties (e.g., boiling point) of networks (e.g., molecules or web graphs)
- Problem: To sample huge sets of structural data statistically (exhaustively generated networks) and calculate the sensitivity of such network descriptors

Example - Structural Interpretation

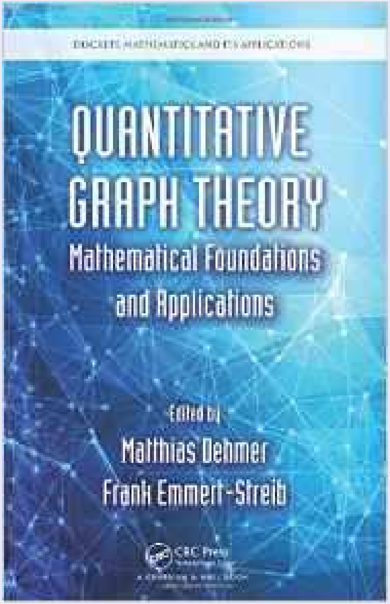
Graph Entropy: $I_f(G) = -\sum_{i=1}^{|V|} \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \log \left(\frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \right)$ where

$f(v_i) := \alpha^{c_1|S_1(v_i,G)|+c_2|S_2(v_i,G)|+\cdots+c_{\rho(G)}|S_{\rho(G)}(v_i,G)|}$ and $c_k > 0, 1 \leq k \leq \rho(G), \alpha > 0$





First Book on Quantitative Graph Theory



Part II

Quantitative Graph Analysis: Problems

Sources of Problems - Structural Graph Measures

- Descriptive approaches for analyzing graphs are often not applicable when analzing graphs
- Therefore, quantitative methods are needed (i.e., graph measures)

Some Problems:

- Often difficult to interpret
- Often difficult to compute (e.g., measures which are based on the automorphism group)
- Sensitivity (i.e., small changes in a graph should result in small changes of the measured value)
- Degeneracy (i.e., non-isomorphic graphs cannot be distinguished)

Uniqueness (Discrimination Power or Degeneracy) of Structural Graph Measures

Definition

Let $I : \mathcal{G} \longrightarrow R$ be a structural descriptor. The uniqueness (discrimination power) of I relates to the ability to discriminate non-isomorphic graphs structurally.

Remark

The degree of the degeneracy can be measured by several quantities (Konstantinova, 1996; Todeschini 1992 etc.), for example

$$S(I) := \frac{|\mathcal{G}| - ndv}{|\mathcal{G}|}.$$

Uniqueness of Structural Graph Measures

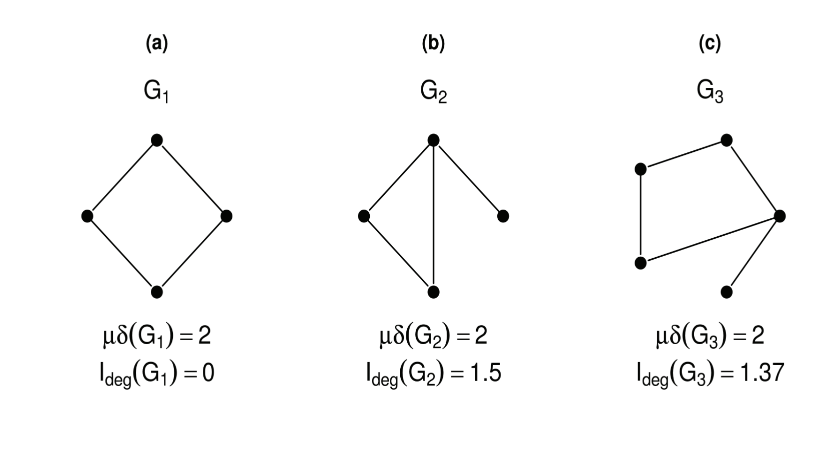
Definition

Calculate I for all $G \in \mathcal{G}$. If $ndv = 0$, then all $G \in \mathcal{G}$ must be non-isomorphic. In this case, we call I complete for the set $\mathcal{G}$.

- So far, no complete graph invariants (structural graph measures) have been found for general graphs.
- Hence, it is clear that any structural graph measure has a certain degree of degeneracy
- Problem: Can we find groups of measures which are highly unique for general graphs?
- Does such a measure only exist for special graph classes (e.g., isomeric structures, alkane trees etc.) ?

Example - Sensitivity

Let $\mu\delta(G) := \frac{\sum_i \delta_i}{N}$ and let $I_{deg}(G) := -\sum_{i=1}^{k} \frac{|\delta_i|}{N} \log \frac{|\delta_i|}{N}$:



(Müller et al. 2011)

Part III

Big Data Sampling Problem

Large Scale Phenomenon

- To perform the study Dehmer et al., we applied Balaban J, Variable Zagreb index, ABC index and various graph entropies to exhaustively generated non-isomorphic graphs.
  - We used exhaustively generated non-isomorphic, connected and unweighted graphs having 9 and 10 vertices. $|N_9| = 261080$ and $|N_{10}| = 11716571$!
  - To generate the networks exhaustively, we have used the `Nauty` package due to McKay (see McKay, 2010)
  - Also, we used exhaustively generated isomers and chemical alkane trees
  - Important question: How strong is the dependency between the uniqueness of I and the $|\mathcal{G}|$
  - To tackle this problem, we performed a statistical analysis

Exemplaric Numerical Results by Using N_{10}

- ABC index: ndv= 11539714 and $S(ABC) = 0,015095$
- Variable Zagreb index (VZI): ndv= 11704386 and $S(VZI) = 0,001040$
- Balaban J index : ndv= 11704386 and $S(J) = 0,001040$
- Magnitude-based information index I_D index : ndv= 11716339 and $S(I_D) = 0,000020$
- Degree-Degree Association index ($I^{\lambda}_{f^{\Delta}_{exp}}$): ndv= 609204 and $S(I^{\lambda}_{f^{\Delta}_{exp}}) = 0,948005$
- Estrada index (EE): ndv= 60054 and $S(EE) = 0,875386$

Part IV

Are new Measures Useful?

Graph Polynomials

Definition

A graph polynomial is a polynomial whose coefficients are defined based on graph invariants.

Examples:

- The Wiener polynomial (also called Hosoya polynomial) has been defined by
$$W_G(z) := \sum_{i=1}^{\rho(G)} d(G,i)z^i.$$
$\rho(G)$ is the diameter of $G = (V, E)$ and $d(G, i)$ is the number of pairs of G having distance i, $d(G, 1) = |E|$.
- Characteristic polynomial $P^c_G(z) := \det(A - zE)$ or distance polynomial $P^d_G(z) := \det(D - zE)$. A is the adjacency matrix and D the distance matrix of G.

A new Non-Standard-Idea:

Instead of using the determinant, we use the permanent of a Matrix A and define the permanental polynomial:

$$P_{\mathrm{per}}^{M(G)}(z) := \mathrm{per}(zE - M(G)) = \sum_{i=0}^{|V|} a_i z^i = 0.$$

We define:

$$I_1^{M(G)}(G) := |z_1^{M(G)}| + |z_2^{M(G)}| + \cdots + |z_k^{M(G)}|$$

,

$$I_2^{M(G)}(G) := \sqrt{|z_1^{M(G)}|} + \sqrt{|z_2^{M(G)}|} + \cdots + \sqrt{|z_k^{M(G)}|}$$

,

$$I_3^{M(G)}(G) := |z_1^{M(G)}| \log(|z_1^{M(G)}|) + |z_2^{M(G)}| \log(|z_2^{M(G)}|) + \cdots + |z_k^{M(G)}| \log(|z_k^{M(G)}|)$$

Discrimination Power of the New Measures

Descriptors →		$I_1^{M(T)}$		$I_2^{M(T)}$		$I_3^{M(T)}$	
Tree classes	$\|T_i\|$	ndv	S	ndv	S	ndv	S
T_{12}	551	119	0.78403	119	0.78403	119	0.78403
T_{13}	1301	417	0.67948	415	0.68101	415	0.68101
T_{14}	3159	828	0.73789	826	0.73852	826	0.73852
T_{15}	7741	2472	0.68066	2470	0.68092	2470	0.68092
T_{16}	19320	5256	0.72795	5246	0.72847	5246	0.72847
T_{17}	48629	14947	0.69263	14944	0.69269	14944	0.69269
T_{18}	123867	32364	0.73872	32347	0.73886	32347	0.73886

Descriptors →		$I_1^{M(G)}$		$I_2^{M(G)}$		$I_3^{M(G)}$	
Graph classes	$\|N_i\|$	ndv	S	ndv	S	ndv	S
N_5	21	0	1.00000	0	1.00000	0	1.00000
N_6	112	2	0.98214	2	0.98214	6	0.94643
N_7	853	0	1.00000	0	1.00000	2	0.99766
N_8	11117	102	0.99082	102	0.99082	109	0.99020
N_9	261080	630	0.99759	624	0.99761	652	0.99750

Part V

Summary, Extensions and Future Work

Summary: Theoretical Aspects

- Sampling structural data on a large scale has been intricate
- Big Data processing becomes a real challenge here
- For this, meaningful and efficient methods are needed
- All structural graph measures have a certain kind of degeneracy
- Most of the measures are highly degenerate. Only a few measures possess high discrimination power
- The discrimination power depends on the graph class
- Entropy-based measures often have high uniqueness. Particularly, this holds for partition-independent measures
- Can structural graph measures help to solve real Big Data problems in data analysis?

Applications and Future Work

- Application of structural graph measures to e.g., financial networks, command and control, communication, and surveillance networks.
- Selection of interesting data sets (data means power!)
- Careful analysis of application areas
- Theoretical Work:
  - Interrelations between graph measures
  - Interrelations between graph distance or similarity measures
  - Statistical analysis