From Graph Theory to Modern Network Science

Miguel Lopes Martins

Critical TW / INESC TEC & CRACS

Data Science Portugal Presentation

Why Graphs Are Ubiquitous

March 4, 2020

Syllabus

1. Introduction
   - Who am I?
   - Genesis of graph theory and definitions
   - Is the definition of Network necessary?
2. Finding Patterns
   - The Subgraph Census Problem
   - Network Motifs
   - Graphlet Degree Distributions
   - Use Cases
3. How to find these Patterns
   - Enumerating Subgraphs
   - Symmetry Breaking
4. Concluding Remarks
   - Useful resources and tools
   - Wrap up and Q&A

Who am I?

- A Data Scientist specially enthusiastic about Deep Learning.
- Personal mission of teaching the computer to see and speak.
- Researcher in Complex Networks and sees graphs everywhere.
- Lover of music, philosophy, drinks and uh... **data** of course.

Genesis of Graph Theory

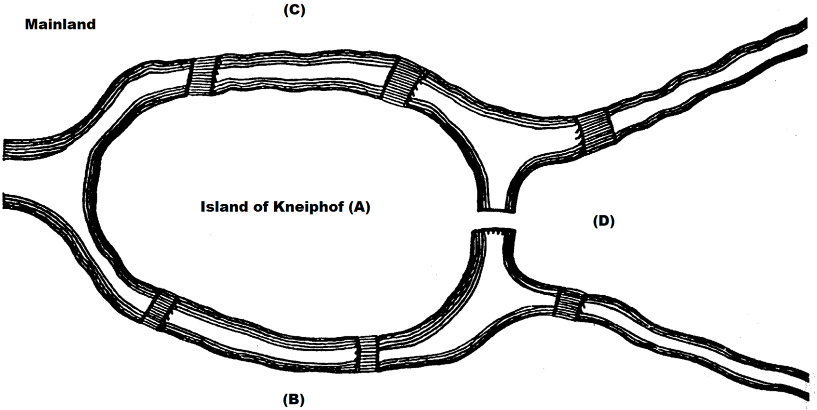


Figure 1: Representation of the diagram of the Seven Bridges of Königsberg presented by Euler in 1735 (Adapted from [1]).

So what is a graph, strictly speaking?

So what is a graph, strictly speaking?

Definition - Graph

A graph $G = (V, E)$, where $V(G)$ is its set of *vertices* that are connected by a set of *edges*, $E(G)$.

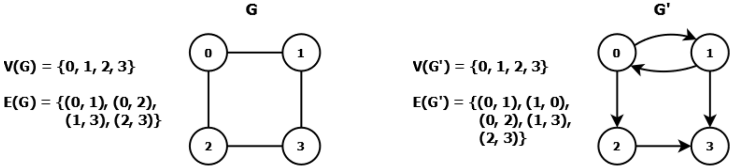


Figure 2: An undirected graph G (left) and a directed graph G' (right). Note how the direction of the edges affects the set of edges of G' .

Genesis of Graph Theory

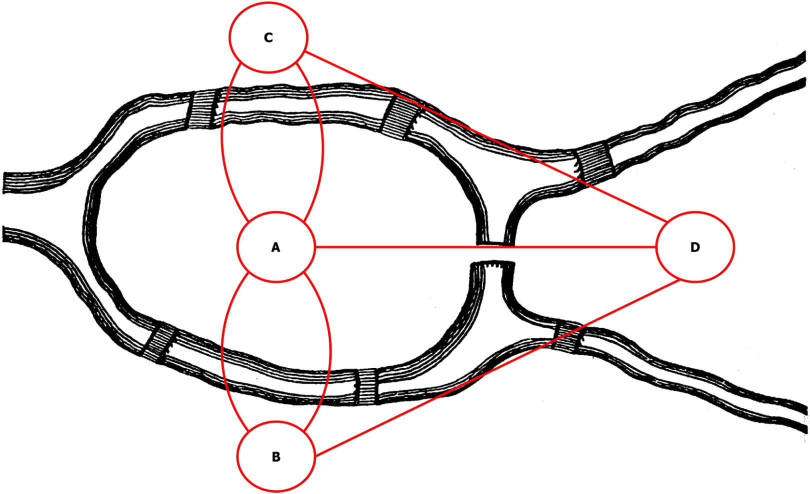
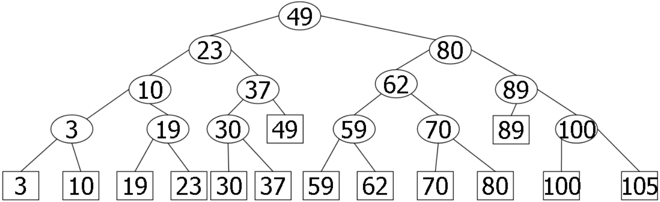


Figure 3: The city of Königsberg represented as a graph.

Spot the Graph



Spot the Graph

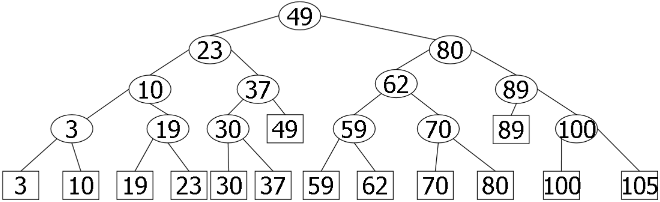
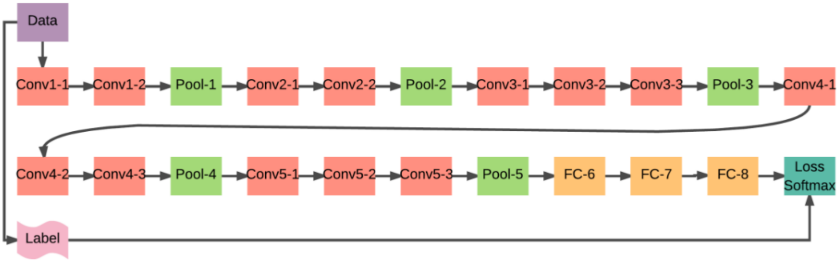


Figure 4: A balanced tree.

Spot the Graph



Spot the Graph

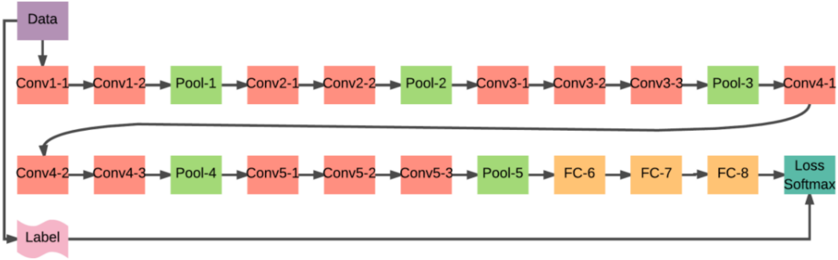
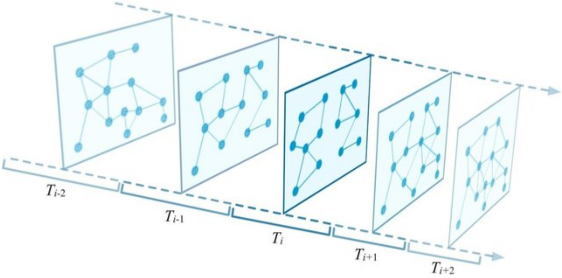


Figure 5: VGG16 Convolutional NN (Source [2])

Spot the Graph



Spot the Graph

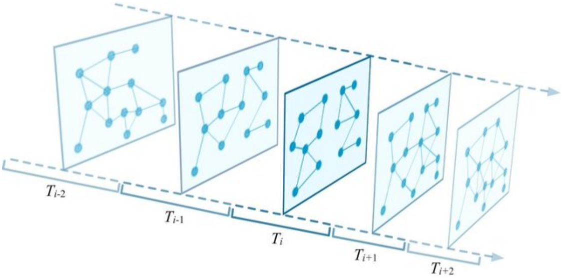
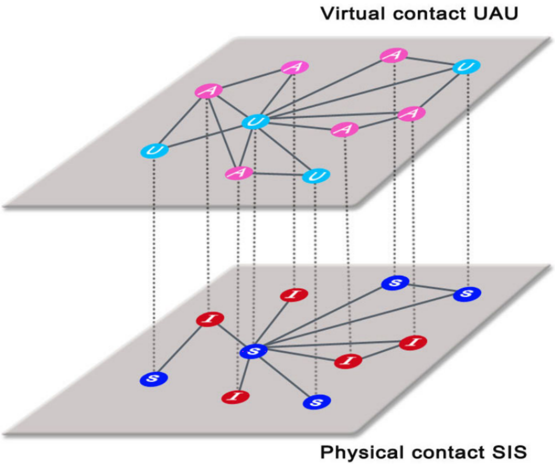


Figure 6: Abstraction of a time-evolving Network (Source [3])

Spot the graph



Spot the graph

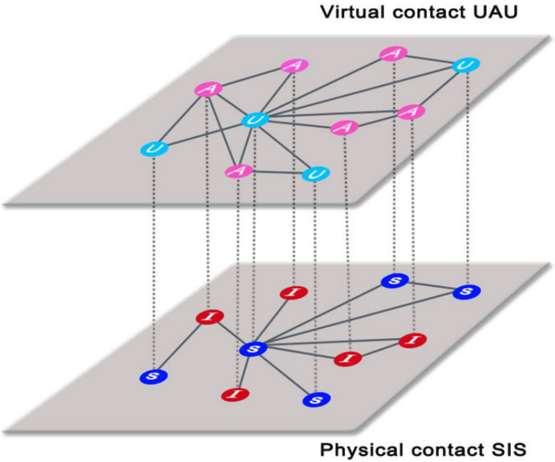


Figure 7: Epidemic multiplex Network (source [4])

Definition - The Subgraph Census Problem

Given some positive integer k and a graph G, count the exact number of distinct occurrences of each of all possible connected induced k-subgraphs of G. Two occurrences are distinct if there is at least one vertex that they do not share.

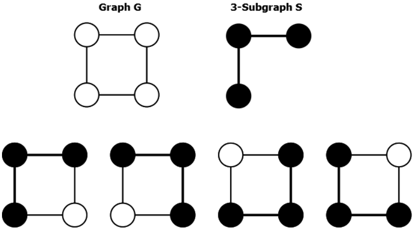


Figure 8: All occurrences of a subgraph of size 3 (right) on graph G (left).

Network Motifs

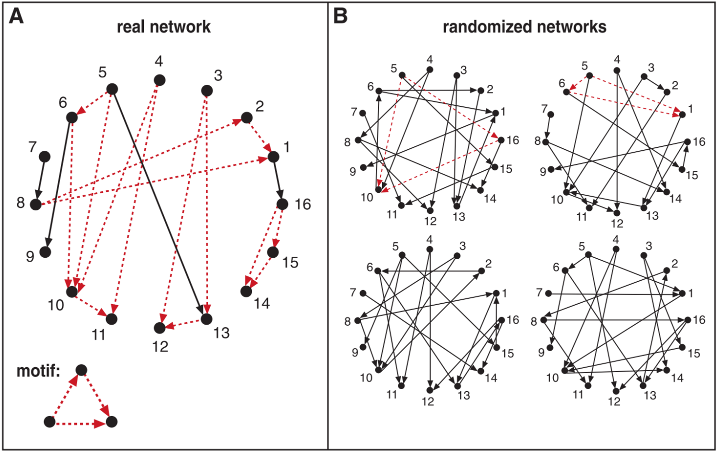


Figure 9: Given a network G a motif M is an induced subgraph that appears more often than expected (source: [5]).

Network Motifs

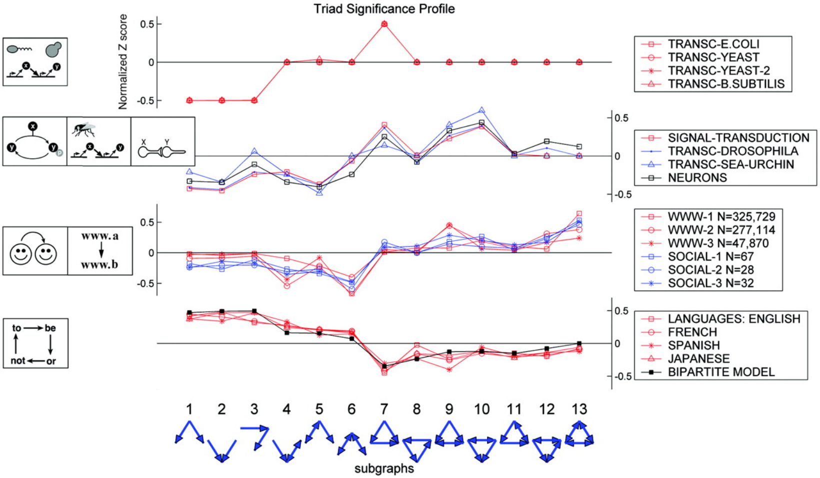


Figure 10: The triad significance profile from from various disciplines (source: [6]).

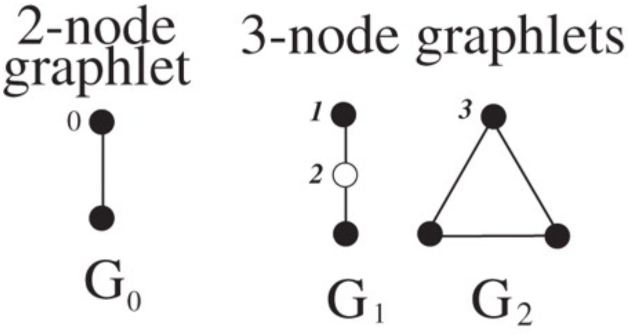


Figure 11: Orbits for all possible graphlets from sizes 2 to 3 (Adapted: [7]).

Graphlet Degree Distributions

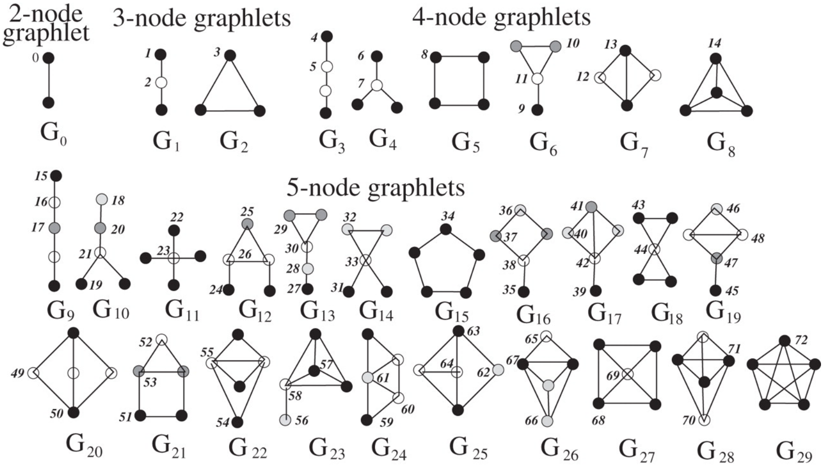


Figure 12: Orbits for all possible graphlets from sizes 2 to 5 (Source: [7]).

Graphlet Degree Distributions

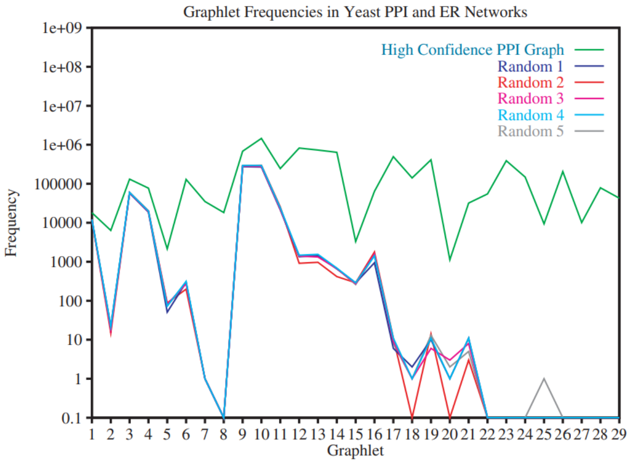


Figure 13: PPI Networks versus Random Networks (Source: [8]).

How to find these patterns efficiently? **ESU**

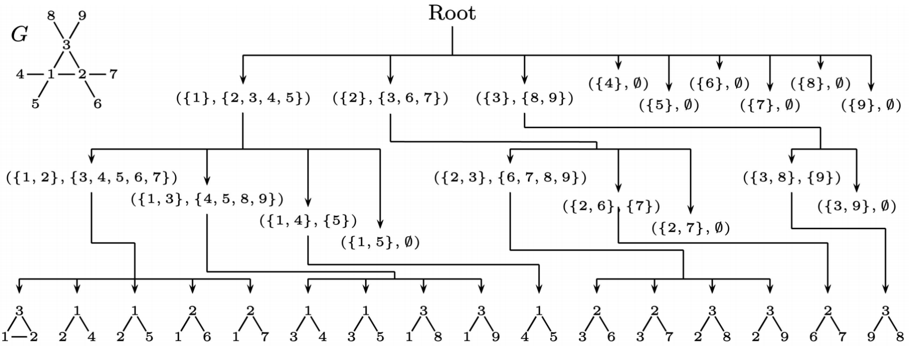


Figure 14: `ESU-TREE` of a graph G (Source: [9]).

Did I Convince you?

Are Graphs/Networks <u>ubiquitous</u> or not?

Useful Resources and tools

- Network Science Book by Barabási.
- NeworkX: a Python Package for Network Analysis.
- geffy: an open-source Network visualization tool.
- gTrieScanner: original code for `G-Tries`.
- Check `FasE` [10], an algortihm which offers a general framework using an adaptive version of `G-Tries`.
- Condensation Decondensation Framework: a POC extending these algorithms by yours truly.
- All the references in this presentation.
- Any questions, feel free to contact me at: `miguelopesmartins@gmail.com`

That's all Folks!

Thank you for your time :)

References I

I. Gribkovskaia, Ø. Halskau Sr, and G. Laporte, "The bridges of königsberg—a historical perspective," *Networks: An International Journal*, vol. 49, no. 3, pp. 199–203, 2007.

H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-vgg16 cnn model for big data places image recognition," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2018, pp. 169–175.

S. Saganowski, "Analysis of social group dynamics," *arXiv preprint arXiv:1708.02000*, 2017.

C. Granell, S. Gómez, and A. Arenas, "Dynamical interplay between awareness and epidemic spreading in multiplex networks," *Physical review letters*, vol. 111, no. 12, p. 128 701, 2013.

References II

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks.," *Science*, vol. 298, no. 5594, pp. 824–827, 2002, ISSN: 1095-9203.

R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, "Superfamilies of evolved and designed networks," *Science*, vol. 303, no. 5663, pp. 1538–1542, Mar. 2004.

N. Pržulj, "Biological network comparison using graphlet degree distribution," *Bioinformatics*, vol. 23, no. 2, e177–e183, 2007.

N. Pržulj, D. G. Corneil, and I. Jurisica, "Modeling interactome: Scale-free or geometric?" *Bioinformatics*, vol. 20, no. 18, pp. 3508–3515, 2004.

References III

S. Wernicke and F. Rasche, “Fanmod: A tool for fast network motif detection,” *Bioinformatics*, vol. 22, no. 9, pp. 1152–1153, 2006, ISSN: 1367-4803. DOI: http://dx.doi.org/10.1093/bioinformatics/btl038.

P. Paredes and P. Ribeiro, “Rand-fase: Fast approximate subgraph census,” *Social Network Analysis and Mining*, vol. 5, no. 1, p. 17, 2015.