Confidence Intervals on Networks

Seong Hwang, Vamsi Ithapu, Sathya Ravi

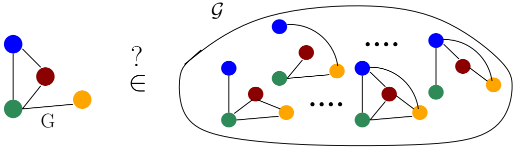
UW Madison

November 9, 2015

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

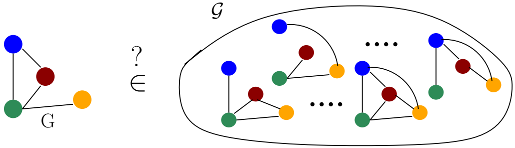
Problems of interest – Motivation

Does G come from $\mathcal{G}$?



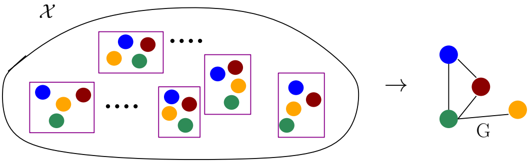
– a testing problem

Does G come from $\mathcal{G}$?



– a testing problem

Given $\mathcal{X}$, estimate G ?



– an estimation problem

Problems of interest – the Question

We need to know/say something more than just G itself

Problems of interest – the Question

We need to know/say something more than just G itself

e.g., both problems above *need* some information about $\mathbb{E}(G)$, $\mathbb{V}(G)$ and other moments of $\hat{G}$

Problems of interest – the Question

We need to know/say something more than just G itself

e.g., both problems above *need* some information about $\mathbb{E}(G)$, $\mathbb{V}(G)$ and other moments of $\hat{G}$

i.e., $G \pm \mathbf{c}\delta_{\mathbf{G}}$ is desirable

Problems of interest – the Question

We need to know/say something more than just G itself

e.g., both problems above *need* some information about $\mathbb{E}(G)$, $\mathbb{V}(G)$ and other moments of $\hat{G}$

i.e., $G \pm \mathbf{c}\delta_{\mathbf{G}}$ is desirable

Can we estimate sub-graph confidence intervals ?

- How bad is my estimate of relationships or disease network ?
- Can we infer similar-behaving neighbourhoods or biomarkers ?
- etc..

Example Applications

Referral Networks

- Do similar NPI sub-graphs correspond to similar drug and/or Medicare amounts incurred ?
- What sub-graphs neighbourhoods correspond to same ranges of Medicare amounts across all drugs ?
- ... and can we relate these similar sub-graphs to some demographics (grography, financial status etc.) ?

Example Applications

Referral Networks

- Do similar NPI sub-graphs correspond to similar drug and/or Medicare amounts incurred ?
- What sub-graphs neighbourhoods correspond to same ranges of Medicare amounts across all drugs ?
- ... and can we relate these similar sub-graphs to some demographics (grography, financial status etc.) ?

Alzheimer's Disease

- Calculating power and sample sizes using estimated *networks* of diease markers
- Mapping the disease progression with *confidence regions* of markers

Example Applications

Referral Networks

- Do similar NPI sub-graphs correspond to similar drug and/or Medicare amounts incurred ?
- What sub-graphs neighbourhoods correspond to same ranges of Medicare amounts across all drugs ?
- ... and can we relate these similar sub-graphs to some demographics (grography, financial status etc.) ?

Alzheimer's Disease

- Calculating power and sample sizes using estimated *networks* of diease markers
- Mapping the disease progression with *confidence regions* of markers

Computer Vision

- Interacting features of interest in detecting or classifying some image/scene/activity

First piece – “Metric”

- Given two graphs derived from some procedure, how do we know if they belong to the same class?

First piece – "Metric"

- Given two graphs derived from some procedure, how do we know if they belong to the same class?
- An easier question (or maybe not!): Given a set of graphs in the same class and a single test graph, can we decide if the test graph is also in the same class?

First piece – "Metric"

- Given two graphs derived from some procedure, how do we know if they belong to the same class?
- An easier question (or maybe not!): Given a set of graphs in the same class and a single test graph, can we decide if the test graph is also in the same class?
- To answer these type of questions, it is clear that we need a way to measure distances between graphs!

First piece – “Metric”

- Given two graphs derived from some procedure, how do we know if they belong to the same class?
- An easier question (or maybe not!): Given a set of graphs in the same class and a single test graph, can we decide if the test graph is also in the same class?
- To answer these type of questions, it is clear that we need a way to measure distances between graphs!
- Fortunately, some smart people have come up with a metric for us to use called as Graph Edit Distance (GED).

GED

Given two weighted graphs G, G', we can use the following operations to transform G to G'

- Add or delete nodes.
- Add or Delete edges.

GED

Given two weighted graphs G, G', we can use the following operations to transform G to G'

- Add or delete nodes.
- Add or Delete edges.

GED is defined as the minimum number of such operations needed to transform G to G'.
Doesn't seem so complicated, but...

Related – Breakthrough of the decade!

Graph Isomorphism *Problem*: Given two graphs G_1, G_2, a graph isomorphism between G_1 and G_2 is a bijective mapping $f : V_1 \to V_2$, V_i is the set of vertices of G_i such that the node and edge weights are preserved under f.

Related – Breakthrough of the decade!

Graph Isomorphism *Problem*: Given two graphs G_1, G_2, a graph isomorphism between G_1 and G_2 is a bijective mapping $f : V_1 \rightarrow V_2$, V_i is the set of vertices of G_i such that the node and edge weights are preserved under f.

- "Graph isomorphism is a kind of NP-complete problem and has to work together with constraints and heuristics in practice" – A survey of graph edit distance by Gao et al (2010).

Related – Breakthrough of the decade!

Graph Isomorphism *Problem*: Given two graphs G_1, G_2, a graph isomorphism between G_1 and G_2 is a bijective mapping $f : V_1 \rightarrow V_2$, V_i is the set of vertices of G_i such that the node and edge weights are preserved under f.

- "Graph isomorphism is a kind of NP-complete problem and has to work together with constraints and heuristics in practice" – A survey of graph edit distance by Gao et al (2010).
- A ray of hope on 11/06/2015!– László Babai says that Graph isomorphism problem can be solved in quasipolynomial time!

Related – Breakthrough of the decade!

Graph Isomorphism *Problem*: Given two graphs G_1, G_2, a graph isomorphism between G_1 and G_2 is a bijective mapping $f : V_1 \rightarrow V_2$, V_i is the set of vertices of G_i such that the node and edge weights are preserved under f.

- "Graph isomorphism is a kind of NP-complete problem and has to work together with constraints and heuristics in practice" – A survey of graph edit distance by Gao et al (2010).
- A ray of hope on 11/06/2015!– László Babai says that Graph isomorphism problem can be solved in quasipolynomial time!
- But the algorithm is too complicated so we will use heuristics based on Dijkstra's algorithm to compute GED :).

Second piece – Generating subgraphs

Recap: We sort of know how to compute distances between graphs.

- The set of subgraphs to consider is exponential and hence it is not feasible to go through the set exhaustively.
- We will now see how to construct a set of *meaningful* subgraphs.

Second piece – Generating subgraphs

Recap: We sort of know how to compute distances between graphs.

- The set of subgraphs to consider is exponential and hence it is not feasible to go through the set exhaustively.
- We will now see how to construct a set of *meaningful* subgraphs.
- One way to go will be to use spectral clustering techniques.
  - Find k clusters. Assign each node to its equivalence class.
  - Let the distance between the equivalence classes be the ratio of number of nodes in each class.
  - Vary k to generate as many subgraphs.
- Alternatively, we propose to use Persistent Homology.

Filtrations

One potential drawback of naive spectral clustering is that the number of connected component of the subgraphs is equal to that of the primary graph that we started out with. Secondly, it is hard to encode the topology of the dataset in spectral clustering.

- A filtration of a given graph G is defined as a sequence of graphs $G_1 \subseteq G_2 \subseteq \cdots$.
- We build a filtration as follows (use whiteboard).