ESCAPE: Efficiently Counting All 5-Vertex Subgraphs

Authors: Ali Pinar, C. Seshadhri, Vaidyanathan Vishal
Presented by: Omar Obeya
Goal

- Count 5-vertex subgraphs
  - Exact
  - Scales
New Problem: 21 5-vertex patterns

Figure 1: Connected 4 and 5-vertex patterns
Counting Patterns
Counting Patterns
Counting Patterns
The Basics
Key Idea: Cutting
Key Idea: Cutting

- Not a Cut
Fragments
Shrinkage
Fragments
Shrinkage
Algorithm
Main Lemma

**Lemma 4.** Consider pattern $H$ with cut set $C$. Then,

$$match(H) = \sum_{\sigma \in \text{match}(H|_C)} \prod_{F \in \text{Frag}_C(H)} \deg_F(\sigma)$$

$$- \sum_{H' \in \text{Shrink}_C(H)} \text{numSh}_C(H, H')match(H')$$
Algorithm

- Pattern count in all graph = sum over all the possible cuts count of pattern - total number of shortage.
Algorithm

- Account for automorphisms
Example
Example
Example
Algorithm

- Pattern count = 2*6 + 1*6 - shrinkage
- Shrinkage = atomorphism * occurrence = 2 * 2
- Answer = 18 - 4 = 14
- Accounting for automorphism = 14/2 = 7
Example
Results
### Performance

Comparison only with 4-vertex PGD

| Dataset                  | $|V|$  | $|E|$  | $|T|$  | Runtime in seconds |
|--------------------------|------|------|------|---------------------|
| soc-brightkite           | 56.7K| 426K | 494K | PGD: 1.20, ESC-4: 0.22, ESC-5: 6.54 |
| tech-RL-caida            | 191K | 1.22M| 455K | PGD: 3.21, ESC-4: 0.25, ESC-5: 5.47 |
| flickr                   | 244K | 3.64M| 15.9M| PGD: 809K, ESC-4: 12.9, ESC-5: 961K |
| ia-email-EU-dir          | 265K | 729K | 267K | PGD: 10.6, ESC-4: 0.18, ESC-5: 8.69 |
| ca-coauth-dblp           | 540K | 3.05M| 444M | PGD: 585, ESC-4: 615, ESC-5: 47.4K |
| web-google-dir           | 876K | 8.64M| 13.4M| PGD: 54.5, ESC-4: 2.94, ESC-5: 71.8 |
| tech-as-skitter          | 1.69M| 22.2M| 28.8M| PGD: 1.90K, ESC-4: 20.3, ESC-5: 1.41K |
| web-hudong               | 1.98M| 14.6M| 5.07M| PGD: 9.40K, ESC-4: 13.6, ESC-5: 534 |
| LiveJournal              | 4.84M| 85.7M| 286M | PGD: 25.9K, ESC-4: 538, ESC-5: 37.1K |
Edge Prediction

Prob. for having another edge
Subgraph Prediction

Ratio between patterns
Future Work

- Scaling to 6-vertex subgraphs??!