Computer Science, Texas A&M University

Modeling the Evolution of Degree Correlation in Scale-Free Topology Generators

Xiaoming Wang

Joint work with Xiliang Liu and Dmitri Loguinov

Internet Research Lab
Computer Science Department
Texas A&M University

April 17, 2008

Agenda

- Introduction
  - Background
  - Motivation
- Analyzing GED
- Generic Framework
- Extending GED results
  - PLRG/BA/AB/GLP
- Comparison
- Conclusion

Introduction

- Topology modeling is an inter-disciplinary topic
  - Computer networks, social/biological/physics systems
- Its goal is to explain how real networks have come into being
  - To develop random graph models that capture the properties found in those systems
- In the context of computer networks, these models are useful for performance evaluations
  - Routing delay, resilience, load balancing, etc.

Background – Metrics

- Metrics of interest
  - Main: degree distribution, assortativity and clustering coefficients
  - Auxiliary: diameter, spectrum, rich-club connectivity
- Degree correlation of level k is the joint degree distribution of k adjacent nodes
- Previous work (Mahadevan 2006) demonstrates that up to 3-level degree correlation suffices to characterize most of existing topologies
- Thus, we focus on degree correlation in this talk

Background – Metrics 2

- Assume an undirected graph $G = (V, E)$ of n nodes with degrees d_1, …, d_n
- L1) Degree distribution CDF $F(x) = P(d<x)$
  - where d is the degree of a random node in V
  - Scale-free graphs exhibit power-law degree distributions $F(x) = 1-(\beta/x)^{\alpha}$

    shape $\alpha>1$

    scale
- L2) Assortativity coefficient $r(G)$ is the Pearson correlation coefficient of node degrees of links
  - It indicates the tendency of high degree nodes connecting frequently to other high degree nodes

Background – Metrics 3

- L3) Clustering coefficient $\gamma(G)$ quantifies how likely the neighbors of a node are to be connected

$$\gamma(G) = \frac{\sum_{i \in \mathcal{V} - \mathcal{V}^{(1)}} \gamma_i}{|\mathcal{V} - \mathcal{V}^{(1)}|}$$

The set of nodes with degree no less than 2

  - where γ_i is individual clustering of node i

$$\gamma_i = \frac{T_i}{d_i(d_i - 1)/2}, \quad d_i \geq 2$$

The number of triangles residing on node i

- We study how $r(G)$ and $\gamma(G)$ change as graph size n grows
  - This allows us to evaluate graph models in terms of their two-/three-node correlation

Background – Models

- Graph models
  - GED (Chung 2002)
  - PLRG (Aiello 2000)
  - BA (Barabasi 1999)
  - AB (Albert 2000)
  - GLP (Bu 2002)
  - HOT (Fabrikant 2002)
  - SWT (Jin 2003)
  - WIT (Wang 2006)

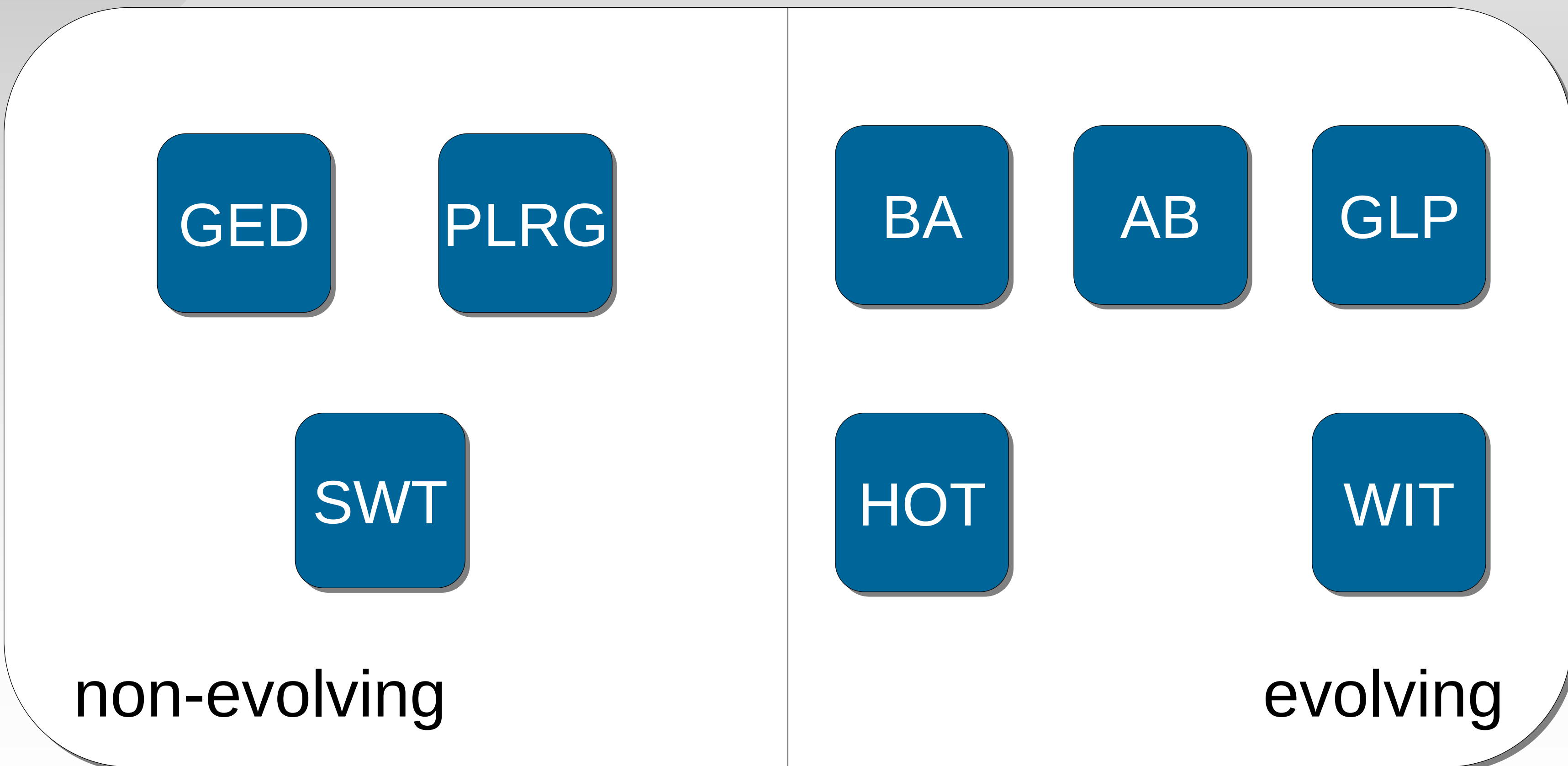
GED, PLRG: Node weight

BA, AB, GLP: Preferential attachment and incremental growth

HOT, SWT, WIT: Geographic distance or wealth evolution and random walk

Background – Models 2

- Classification
  - Non-evolving and evolving



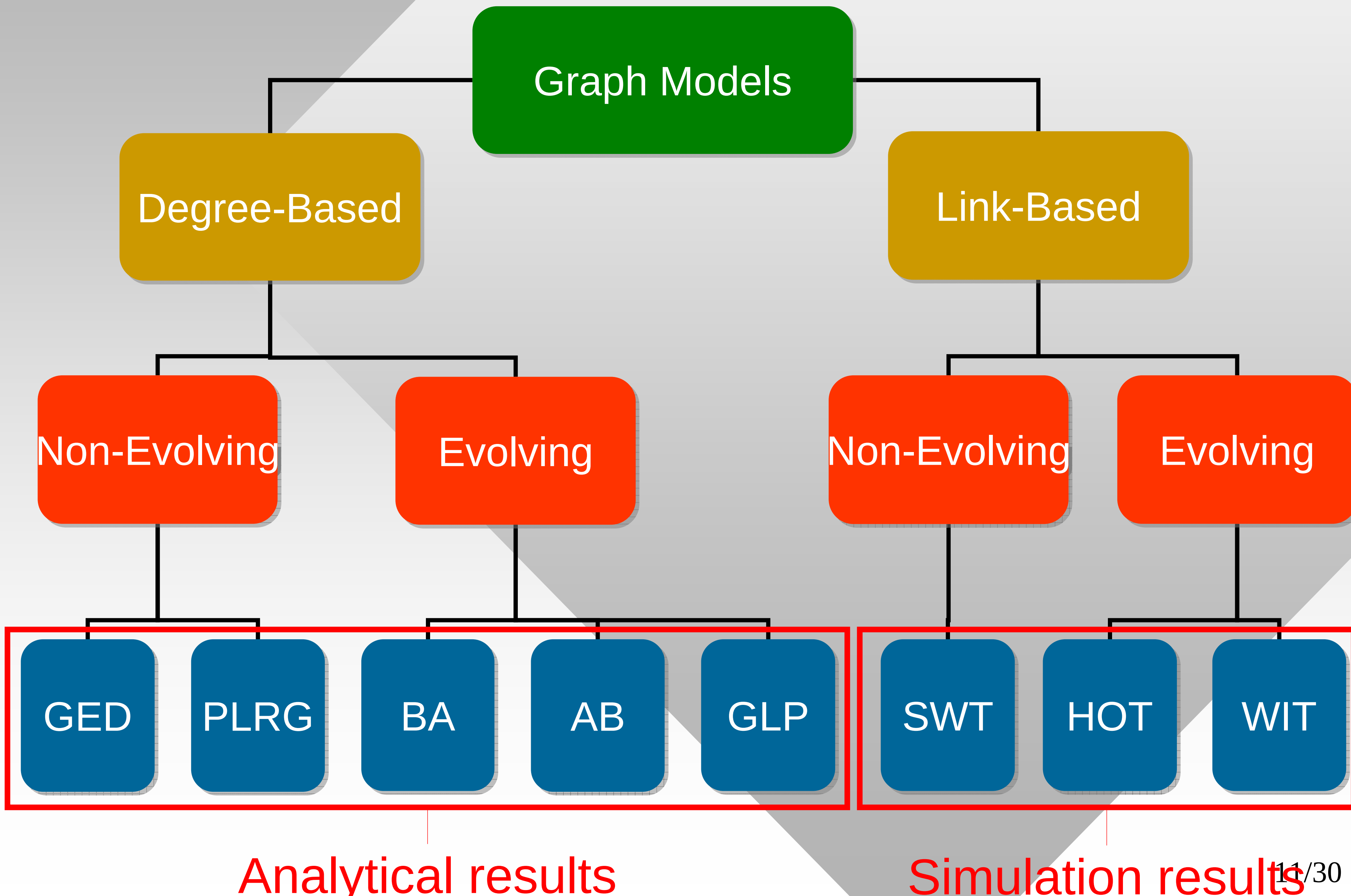
Background – Models 3

- Classification
  - Degree-based and link-based

Motivation

- Many large-scale networks show invariant degree correlation
  - $r(G)$ and $\gamma(G)$ stay constant with growing n
- However, many algorithms have decreasing degree correlation as n increases
- No prior analysis has examined this issue in topology models
  - Only partial results are available in the literature for $\alpha \geq 2$
- Goal is to study $1 < \alpha < 2$ for popular generators

Roadmap


Graph Models
Degree-Based
Link-Based
Non-Evolving
Evolving
Non-Evolving
Evolving
GED
PLRG
BA
AB
GLP
SWT
HOT
WIT
Analytical results
Simulation results

GED – Basics

- GED assigns random weights w_i drawn from Pareto distribution $F(x)$
- It then creates each link with probability p_{ij}

$$p_{ij} = \min\left(\frac{w_i w_j}{D}, 1\right)$$

  - where D is total weight: $D = \sum_{k=1}^{n} w_k$
- Next, we formalize the relationship between node weights and edge-existence probability
  - We keep our formalization general enough so that it can be applied in any graph G

Link Formation – General Discussion

- Define $\pi(x, y)$ to be the probability of two nodes being connected given their weights x and y

$$\pi(x,y) = P(i \leftrightarrow j | w_i = x, w_j = y)$$

- For GED, this π-function is simply given by

$$\pi(x,y) = \min\left(\frac{xy}{D}, 1\right)$$

- We next see how $E[r(G)]$ and $E[\gamma(G)]$ can be computed using the π-function

Assortativity Coefficient - GED

- For any graph, we can express $E[r(G)]$

$$E[r(G)] = \frac{E[d]\rho - E^2[d^2]}{E[d]E[d^3] - E^2[d^2]}$$

$E[d^k]$ is the k-th moment of degree

  - where ρ is given by:

$$\rho = n \iint xy\pi(x,y) f(x) f(y) dx dy$$

Graph size

π-function

Weight PDF

- Theorem 1: the expected assortativity coefficient of GED graphs is asymptotically:

$$E[r(G)] = \begin{cases} \Theta\left(-n^{1-\alpha}\right) & 1 < \alpha < 2 \\ \Theta\left(-n^{-1}\log^2 n\right) & \alpha = 2 \\ 0 & \alpha > 2 \end{cases}$$

Clustering Coefficient

- Theorem 2: With Pareto distributed weights, the expected GED clustering is asymptotically:

$$E[\gamma(G)] = \begin{cases} \Theta(n^{1-\alpha} \log n) & 1 < \alpha < 2 \\ \Theta(n^{-1} \log^2 n) & \alpha = 2 \\ \Theta(n^{-1}) & \alpha > 2 \end{cases}$$

- Derivations are fairly convoluted
  - See the paper for details
- We next extend our results of GED to other degree-based models

Generic Framework – Equivalence

- Assume two topology algorithms A and B
  - With edge-existence probabilities $\pi_A(x, y)$ and $\pi_B(x, y)$, respectively
  - With the same weight distribution $F(x)$
- We say A is asymptotically π-equivalent to B if $\pi_A(x, y)$ is upper/lower bounded by $\pi_B(x, y)$:

$$\boldsymbol{c_L \pi_B(x,y) \leq \pi_A(x,y) \leq c_U \pi_B(x,y)}$$

  - where $0 < c_L < 1$ and $c_U \geq 1$ are some constants

Generic Framework – Equivalence 2

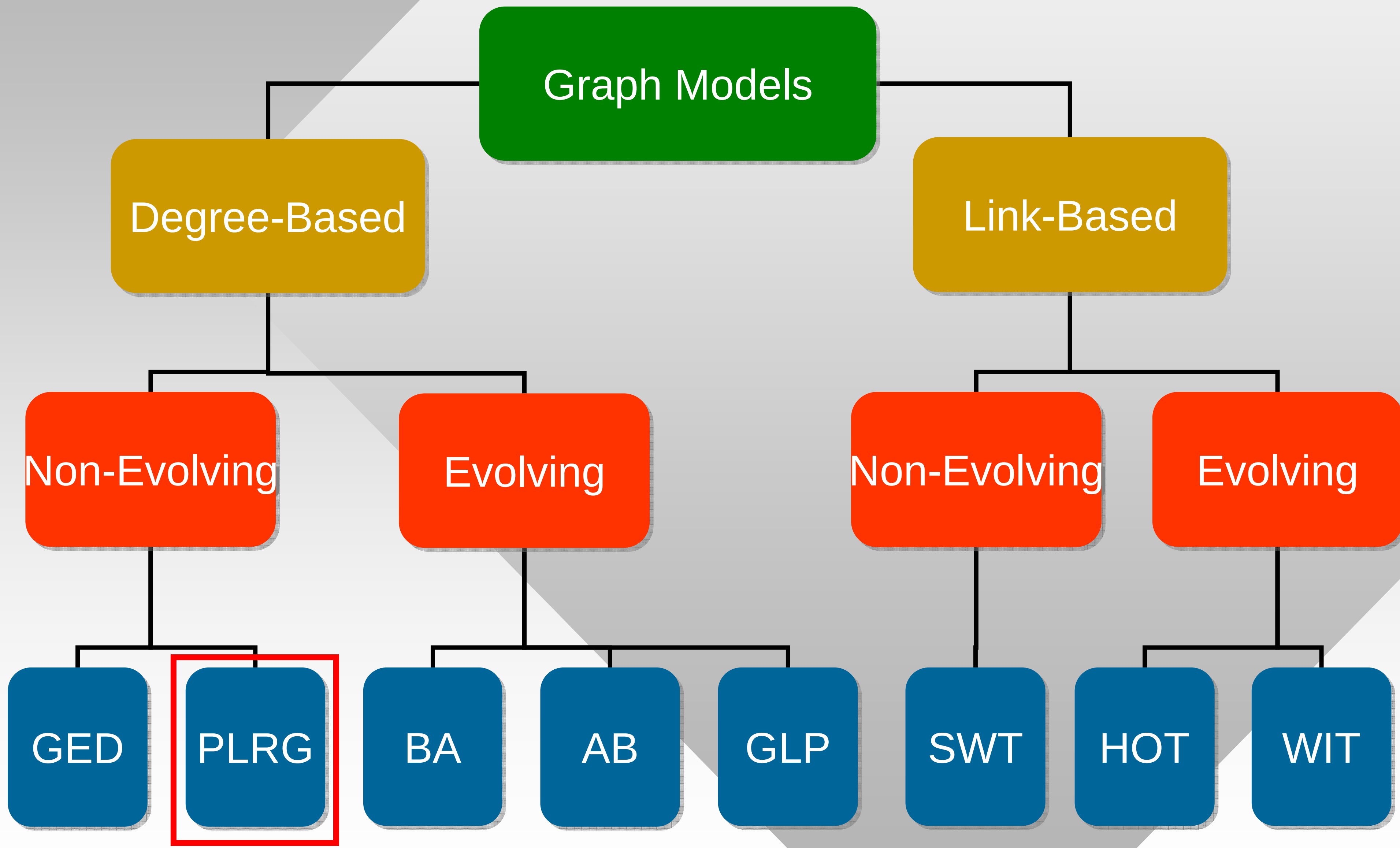
- Properties of π-equivalence
  - Symmetric and transitive
- Theorem 3: If algorithm A is π-equivalent to B, then their expected correlation coefficients have the same asymptotic trends:

$$\lim_{n\to\infty} \frac{E[r(G_A)]}{E[r(G_B)]} = c_r \qquad \lim_{n\to\infty} \frac{E[\gamma(G_A)]}{E[\gamma(G_B)]} = c_\gamma$$

  - where c_r and c_γ are bounded by:

$$c_L \le c_r \le c_U, \qquad \frac{c_L^3}{c_U} \le c_\gamma \le \frac{c_U^3}{c_L}$$

Roadmap



Graph Models
Degree-Based
Link-Based
Non-Evolving
Evolving
Non-Evolving
Evolving
GED
PLRG
BA
AB
GLP
SWT
HOT
WIT

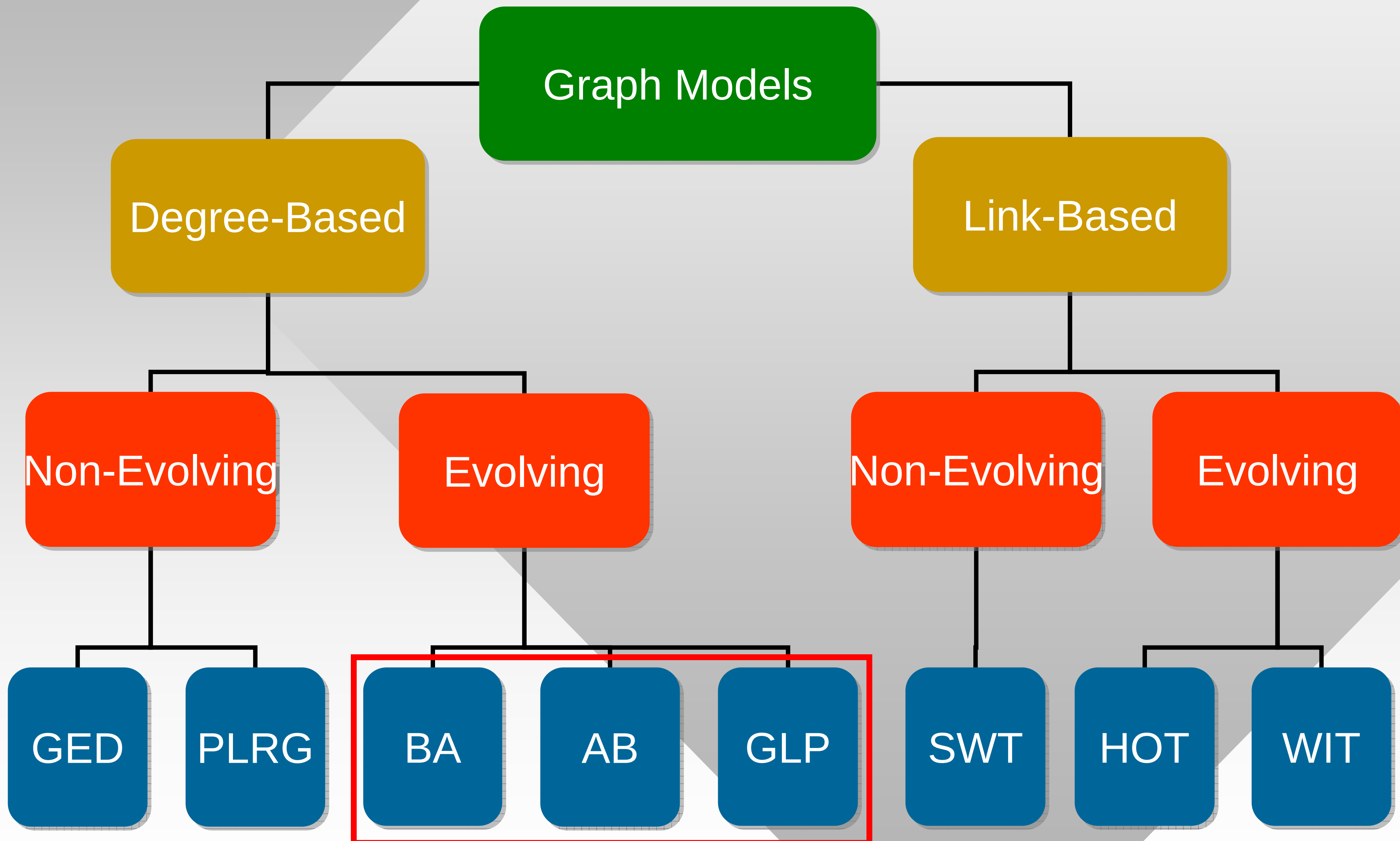
Non-Evolving Model – PLRG

- PLRG generates w_i virtual copies for node i
- Then, it randomly pairs up virtual nodes to form actual links
- Theorem 4: PLRG's π-function is:

$$\pi(x, y) = 1 - e^{-xy/D}$$

- Theorem 5: PLRG is π-equivalent to GED
- Thus, PLRG has the same asymptotic degree correlation as GED

Roadmap



Graph Models
Degree-Based
Link-Based
Non-Evolving
Evolving
Non-Evolving
Evolving
GED
PLRG
BA
AB
GLP
SWT
HOT
WIT

Evolving Model – GLP

- At each time step t,
  - With probability p, GLP adds m new links among existing nodes
  - With probability $1-p$, it adds a new node with m new links connecting to existing nodes
- The probability of selecting existing node i is proportional to its degree $d_i - \lambda$

$$p_i(t) = \frac{d_i(t) - \lambda}{\sum_{k=1}^{n(t)} (d_k(t) - \lambda)}$$

$\lambda \leq 1$ is the shift parameter

- No weights used in GLP construction
  - Thus, we cannot directly apply the GED analysis here

Evolving Model – GLP 2

- However, we can still use the framework by setting weight $w_i = d_i - \lambda$ for each node i

- Theorem 6: GLP's π-function is given by:

$$\pi(x,y) = 1 - 1/2\exp\left(-c_1\frac{xy}{D}\left(p + c_2 x^{\alpha-2}\right)\right) - 1/2\exp\left(-c_1\frac{xy}{D}\left(p + c_2 y^{\alpha-2}\right)\right)$$

Pareto shape parameter

Total weight

  – where c_1 and c_2 are constants

- Theorem 7: GLP is π-equivalent to PLRG
  – By transitivity, GLP is also π-equivalent to GED

Evolving Model – BA/AB

- BA is the same as GLP with $\alpha=2$, $p=1$, and $\lambda=0$
- AB without edge rewiring is GLP with $\lambda = -1$
- Corollary 1: BA and AB without rewiring are π-equivalent to GED
- Consider that BA has $\alpha=2$ and its clustering is

$$E[\gamma(G)] = \Theta(n^{-1} \log^2 n)$$

- This result has been derived (Barabasi 1999)
  - It is just a coproduct of our generic framework

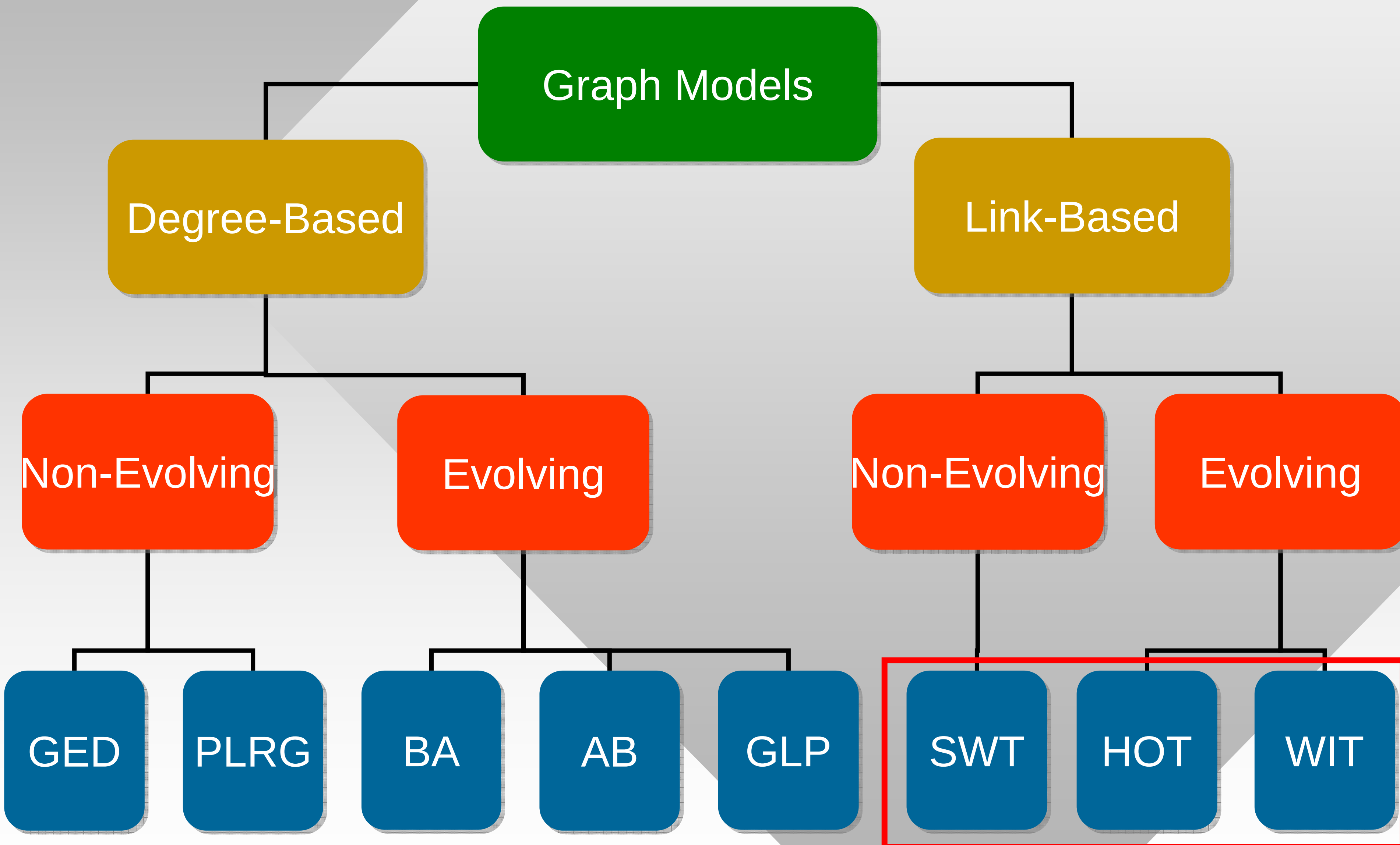
Degree-Based Models – Discussion

- Conclusion: all studied degree-based algorithms become uncorrelated as $n \to \infty$ and their decay rates are given by:

$$E[r(G)] = \begin{cases} \Theta\left(-n^{1-\alpha}\right) & 1 < \alpha < 2 \\ \Theta\left(-n^{-1}\log^2 n\right) & \alpha = 2 \\ 0 & \alpha > 2 \end{cases}$$

$$E[\gamma(G)] = \begin{cases} \Theta(n^{1-\alpha}\log n) & 1 < \alpha < 2 \\ \Theta(n^{-1}\log^2 n) & \alpha = 2 \\ \Theta(n^{-1}) & \alpha > 2 \end{cases}$$

Roadmap



Graph Models
Degree-Based
Link-Based
Non-Evolving
Evolving
Non-Evolving
Evolving
GED
PLRG
BA
AB
GLP
SWT
HOT
WIT

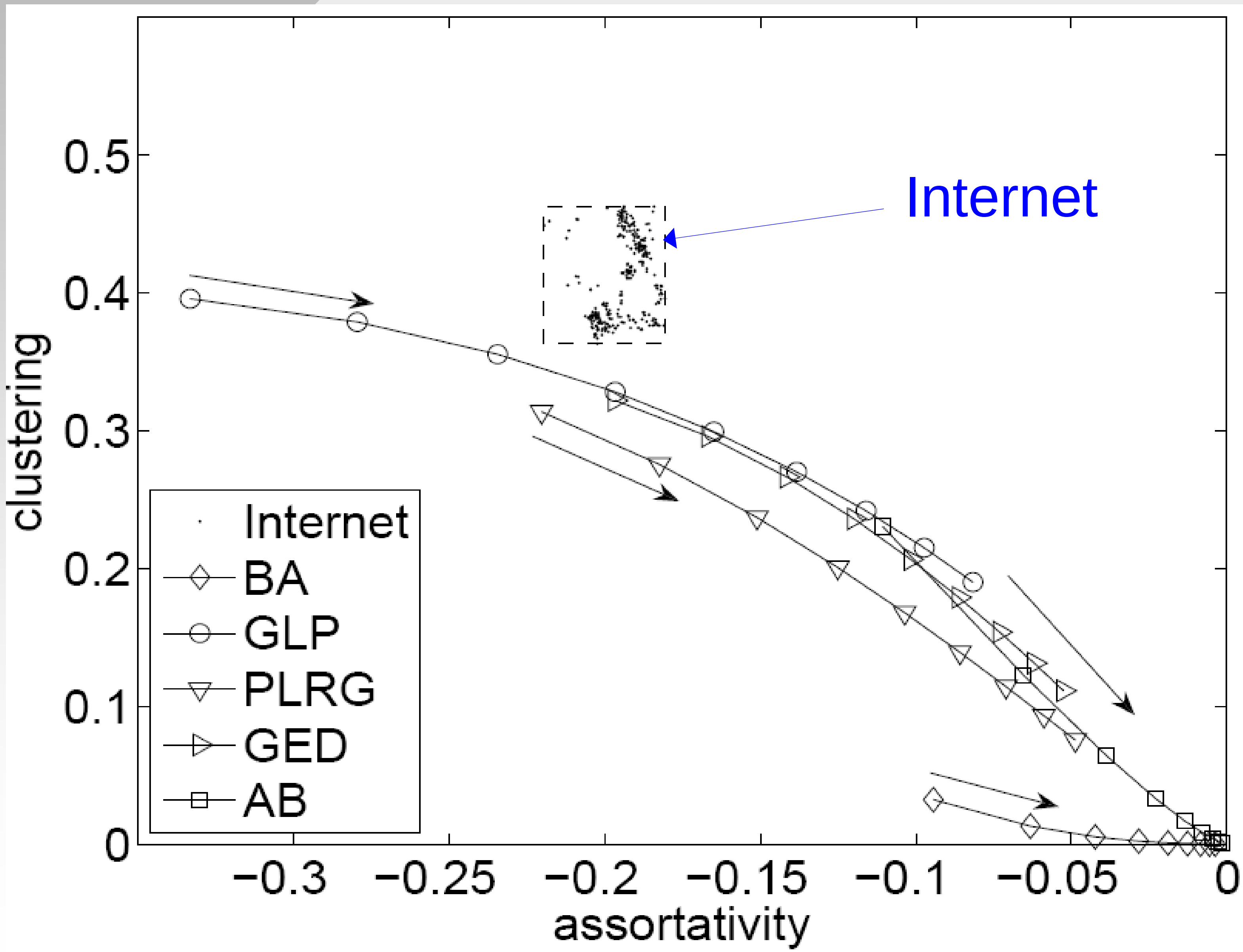
Link-Based Algorithms

- SWT forms p percent of links using geographic preference and the rest using random pairing
- HOT models each attachment decision as an optimization problem with two objectives
  - Last-mile cost and reachability
- WIT adjusts the number of links based on a stochastic wealth process
  - Neighbor selection is based on random walks
- We study these algorithms in simulations

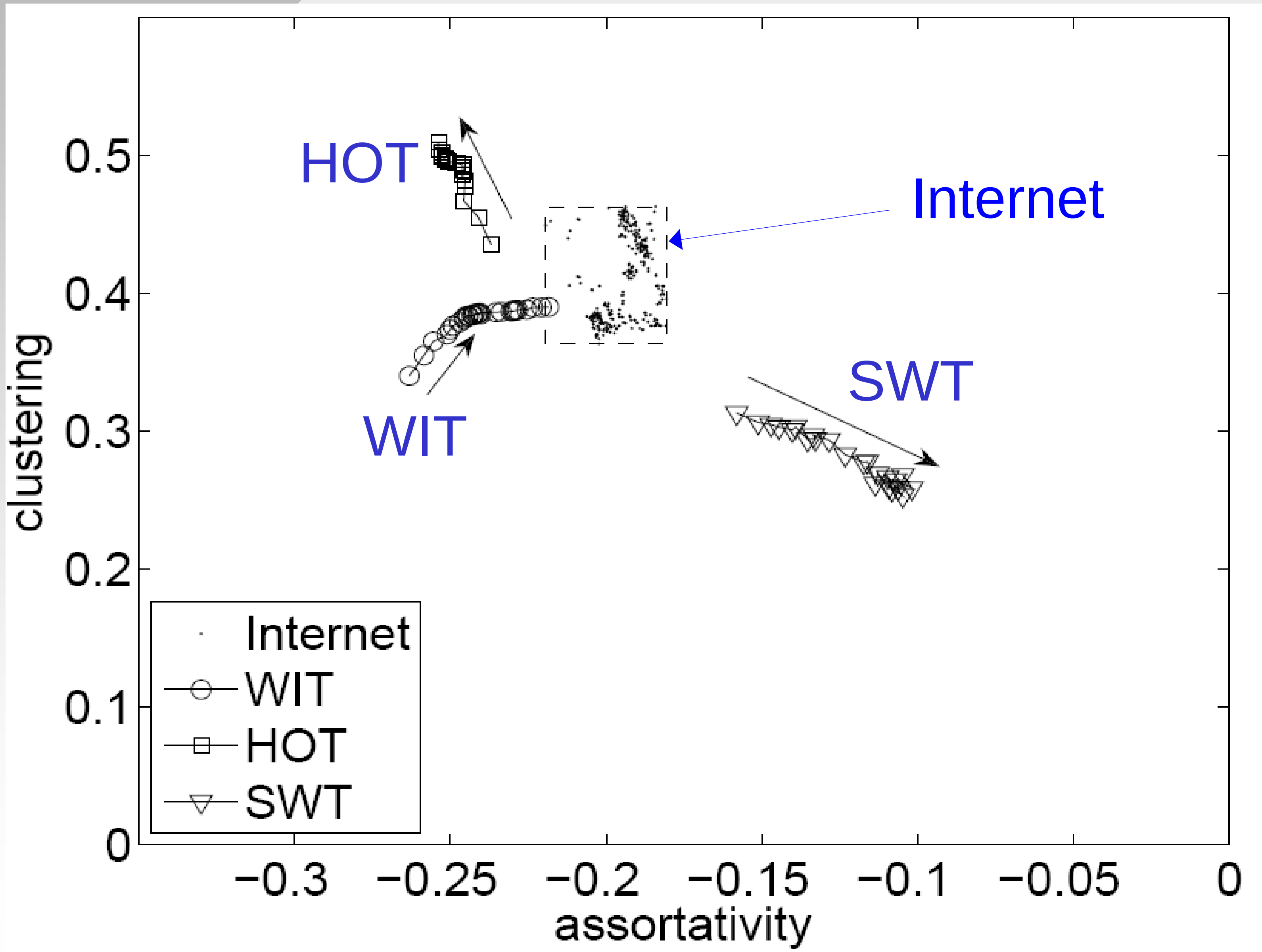
Comparison

- We use the Internet AS-level graph as the benchmark
  - Route-Views and RIPE
- We extract $E[r(G)]$ and $E[\gamma(G)]$ from historical data of the BGP graph observed during the last 7 years
  - The graph size increases from $4{,}000$ to $23{,}000$ nodes
- The data shows $E[r(G)]$ and $E[\gamma(G)]$ of the Internet do not change much over the years
- Next, we compare the studied models

Comparison – Degree-Based



Comparison – Link-Based



Conclusion

- We developed an analytical framework for modeling degree-based algorithms
  - We found that all studied degree-based methods become uncorrelated as $n \to \infty$
- Our simulations showed that some of the studied link-based algorithms were capable of keeping $E[r(G)]$ and $E[\gamma(G)]$ time-invariant
- Future work
  - Extension to other degree-based methods
  - Analysis of link-based models
  - Higher-order degree correlation