

Generating Evolving Property Graphs with Attribute-Aware Preferential Attachment

Amir Aghasadeghi (aa3657@Drexel.edu)

Julia Stoyanovich (stoyanovich@drexel.edu)

College of Computing & Informatics

Drexel University, Philadelphia, PA

DBTest
June 15th, 2018



Goals

In this work

- We make progress towards a data generator for evolving property graphs
- Such graphs represent evolution of
 - Graph topology
 - **Vertex and edge attributes**

Ultimate goal

- Create a standardized benchmark with predefined queries to measure performance of evolving graph databases

Outline

1

Introduction

2

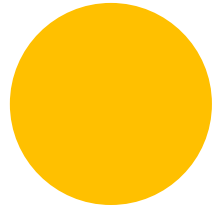
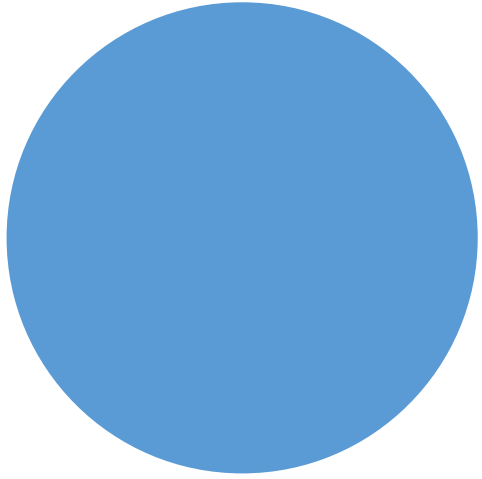
Random Walk
and Recursive
Search

3

Attribute-Based
Preferential
Attachment

4

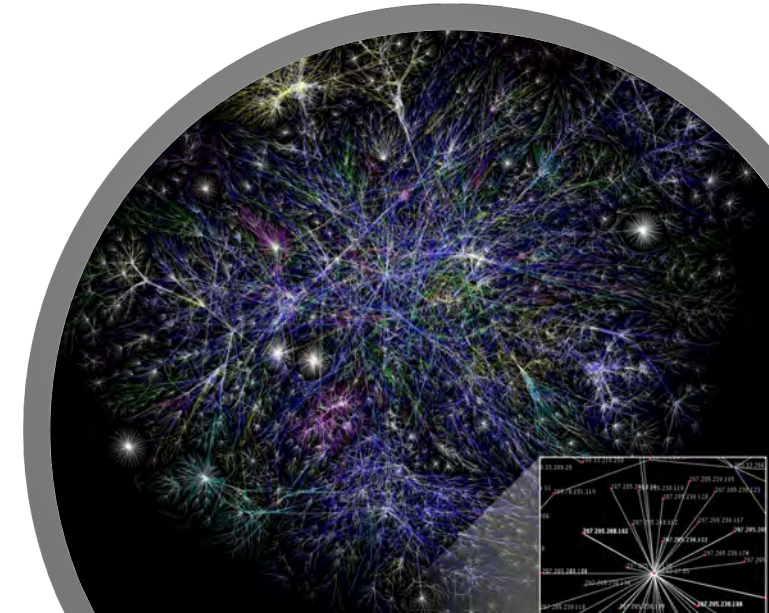
Graph
Generation



Introduction

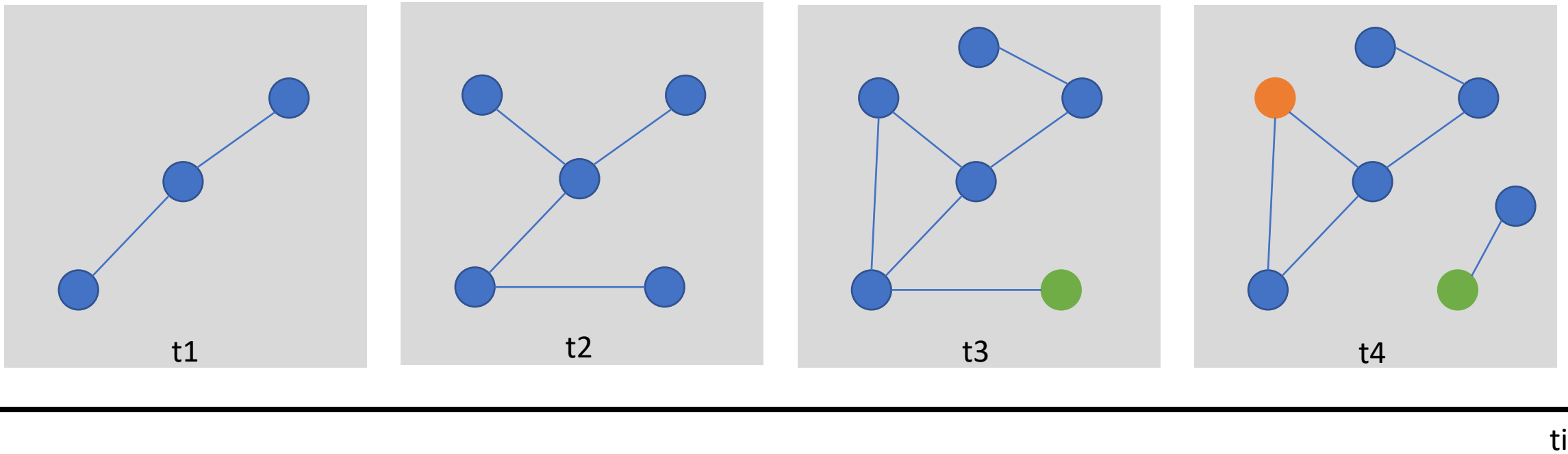
Graphs

Graphs are used to represent a plethora of phenomena

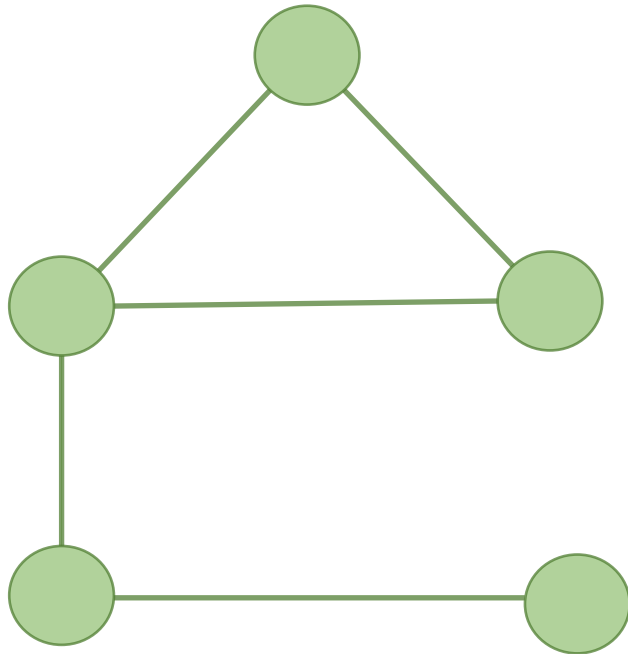


Evolving Graphs

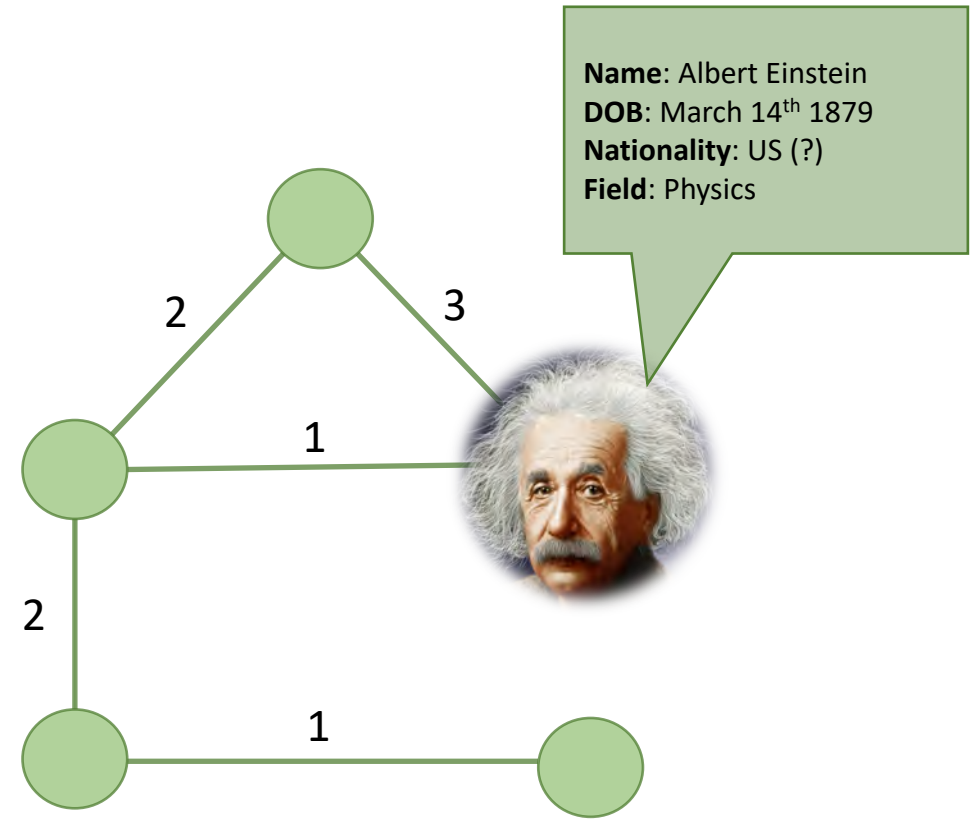
- These phenomena can change over time



Property Graphs



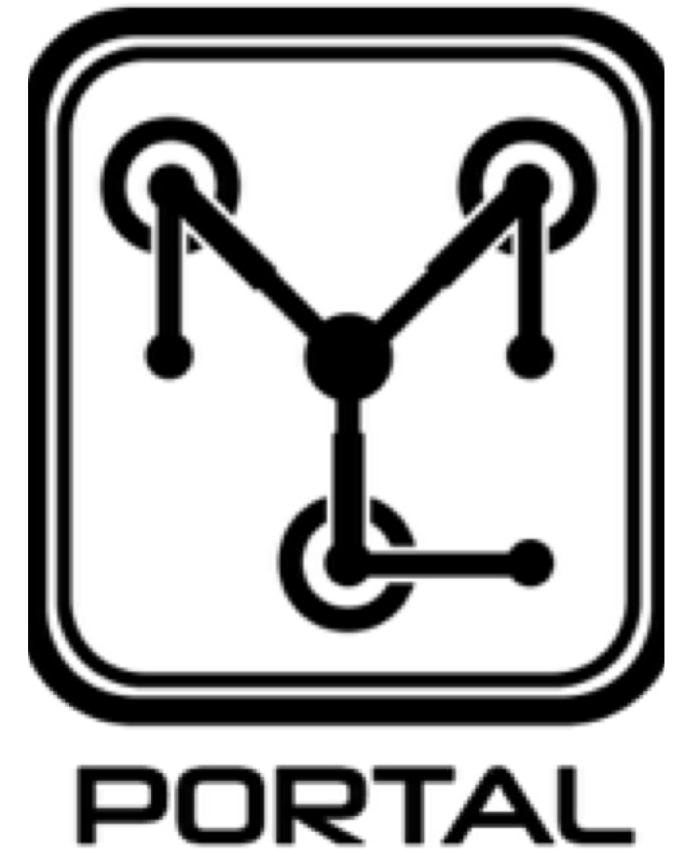
Topology



Attributes

- Need scalable and generalizable systems for evolving graphs
- We built such a system: **Portal**
- Standardized datasets necessary to measure performance
 - Hard to obtain real evolving graph datasets
 - Need to control graph characteristics

We need a model to generate realistic syntactic evolving graphs



Previous Works

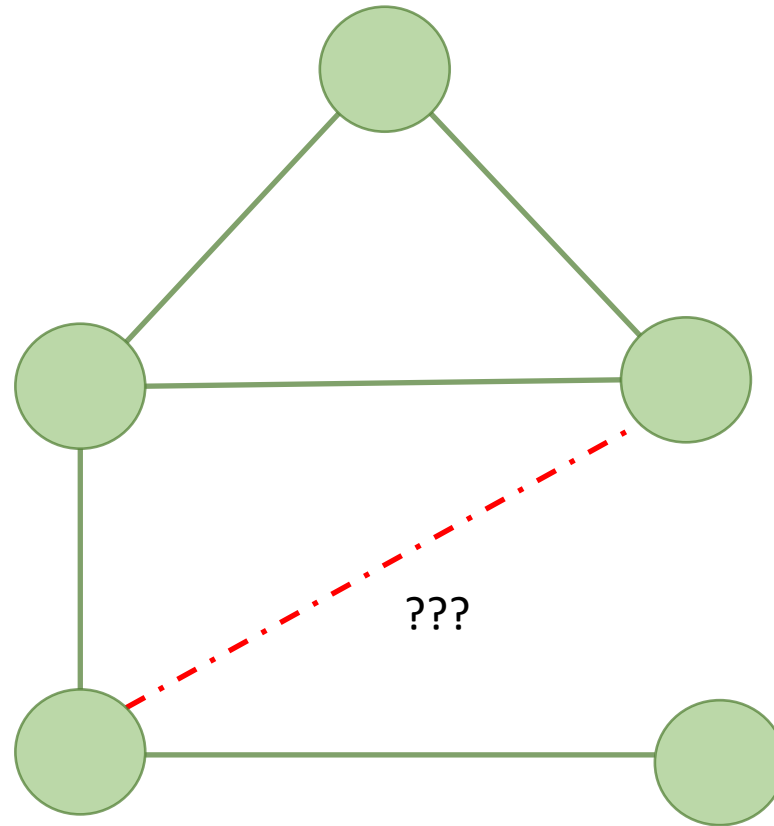
Evolution of graph topology over time

- Random Walk
- Forest Fire
- Solely focus on structural evolution
- No attributes

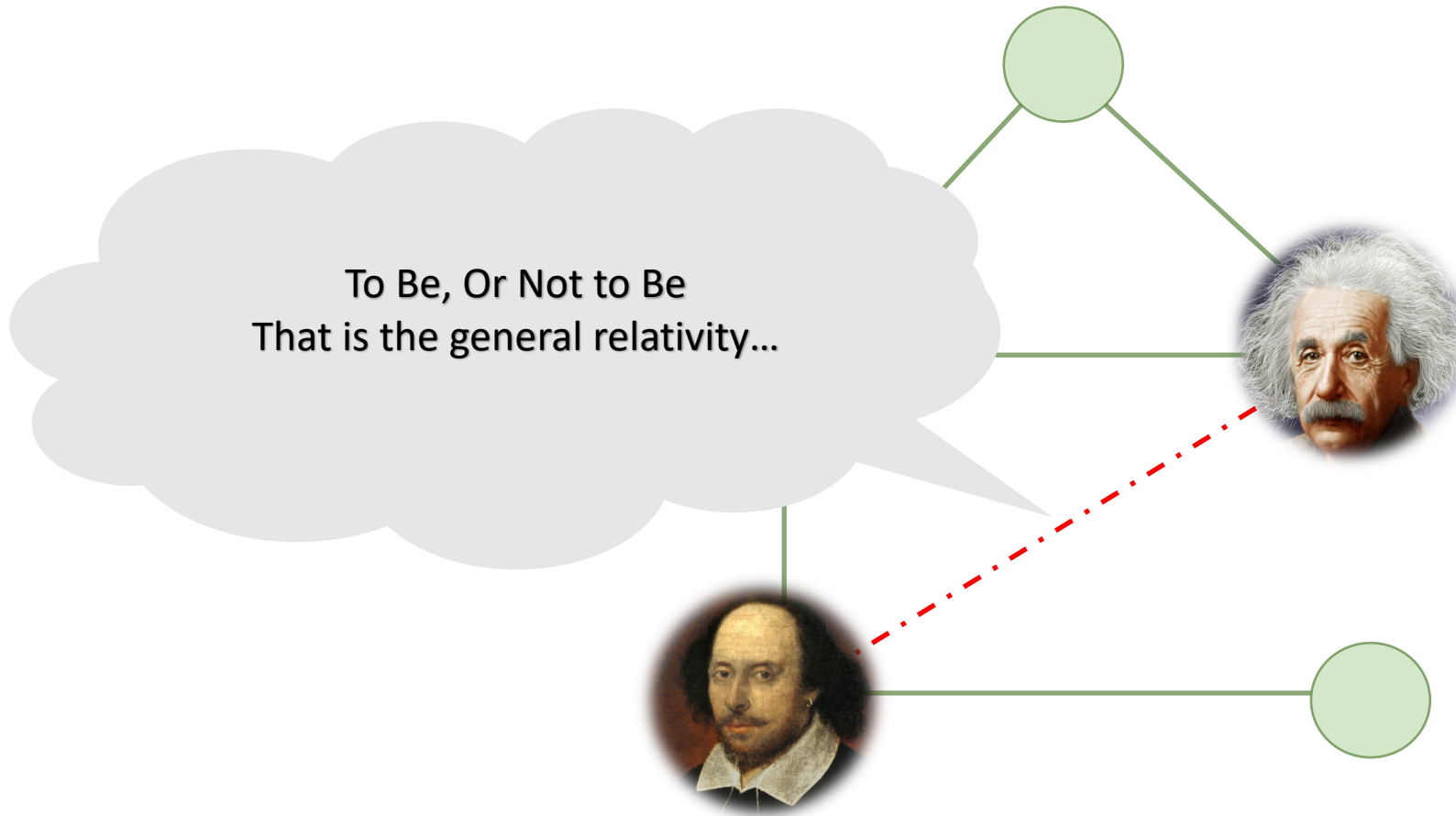
LDBC Social Network Benchmark

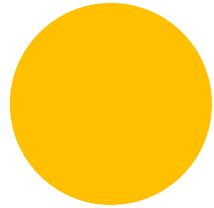
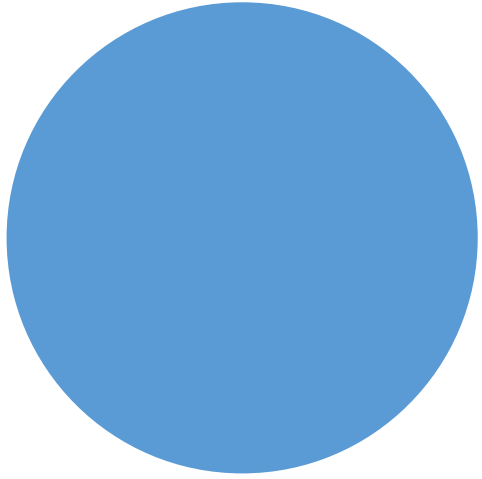
- A realistic synthetic social network
- Has attributes
- Doesn't use a principled evolution model

Attributes ...



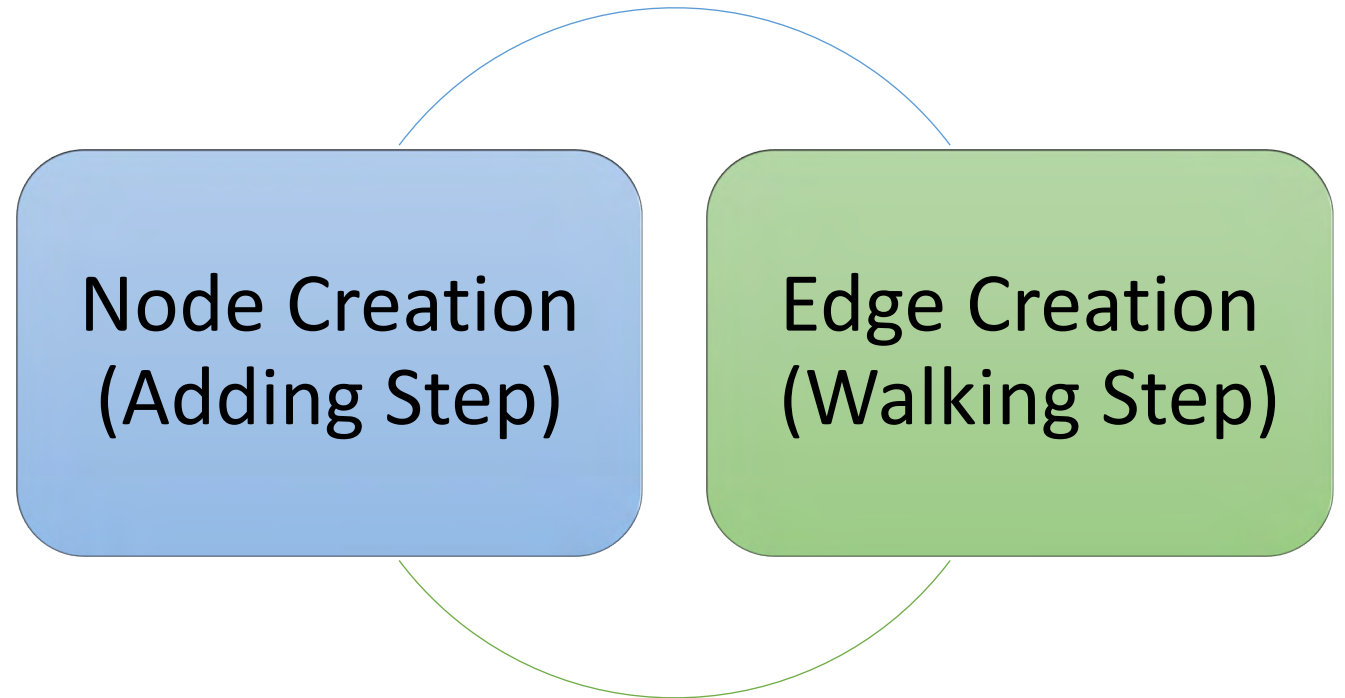
Attributes matter



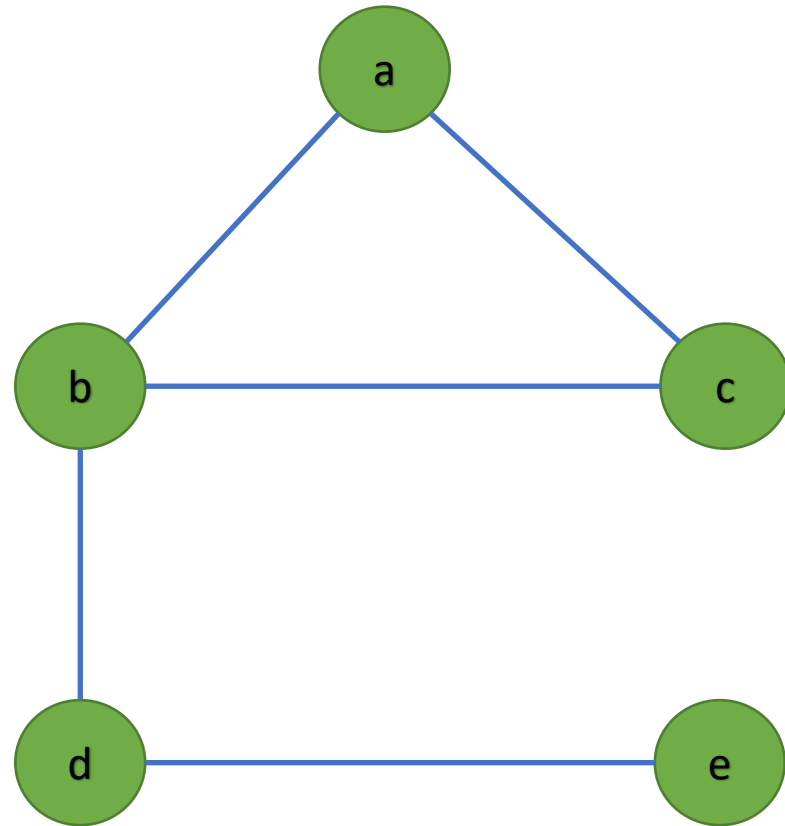


Random Walk and Recursive Search

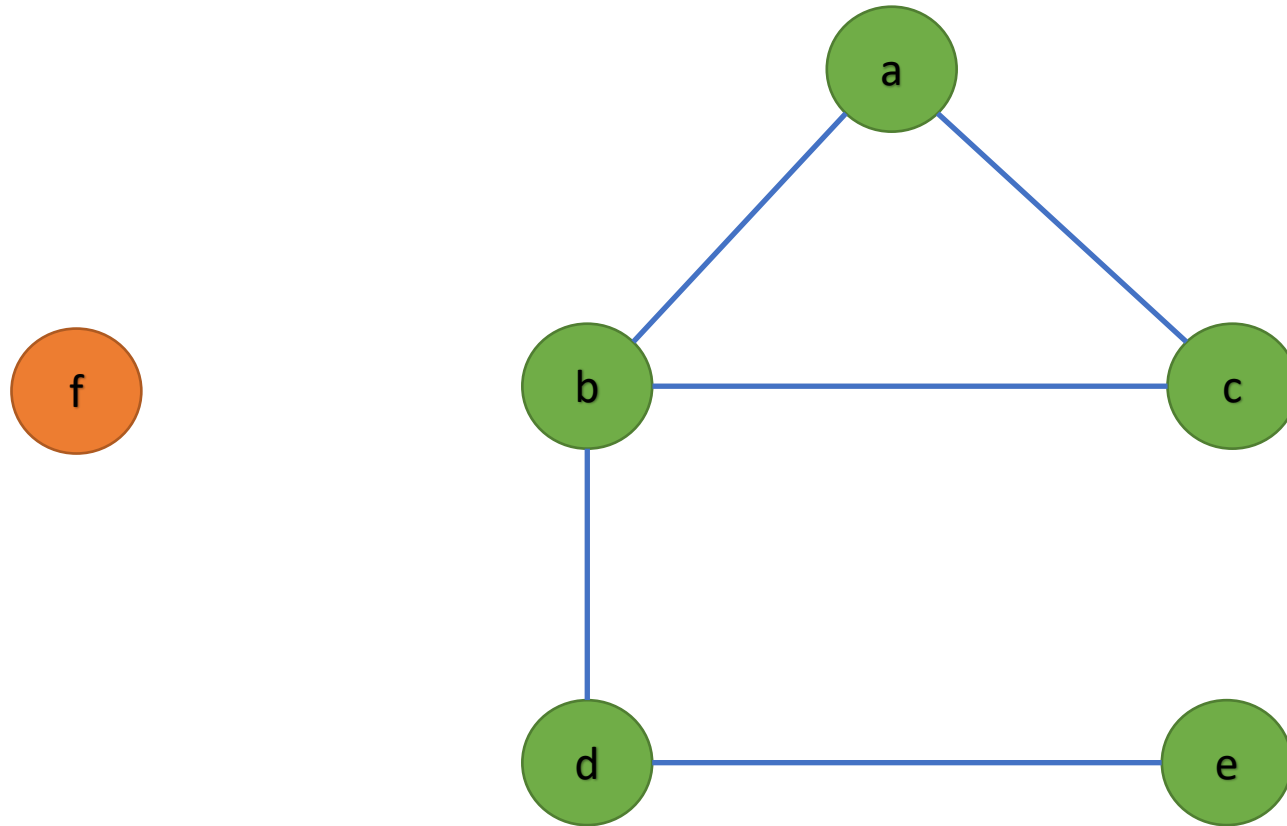
Random Walk (RW)



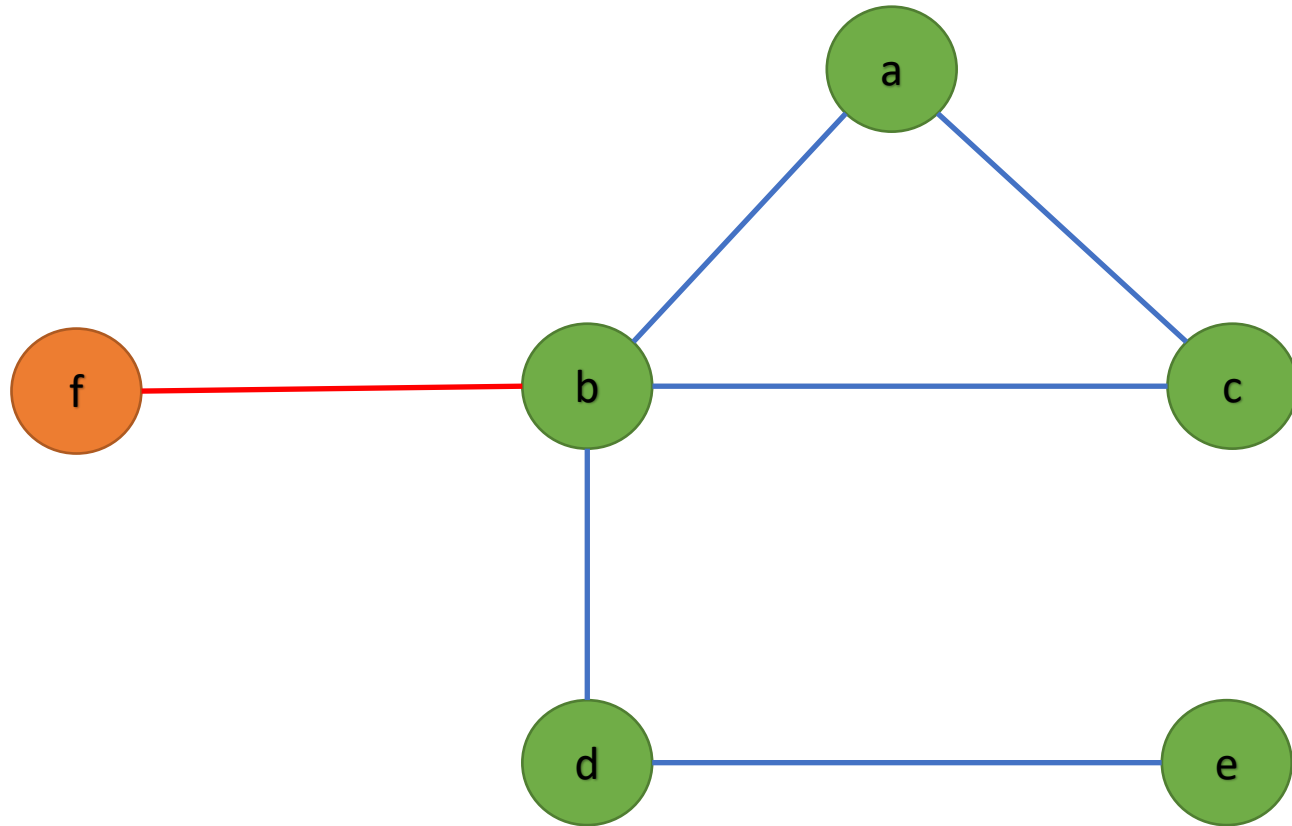
Random Walk



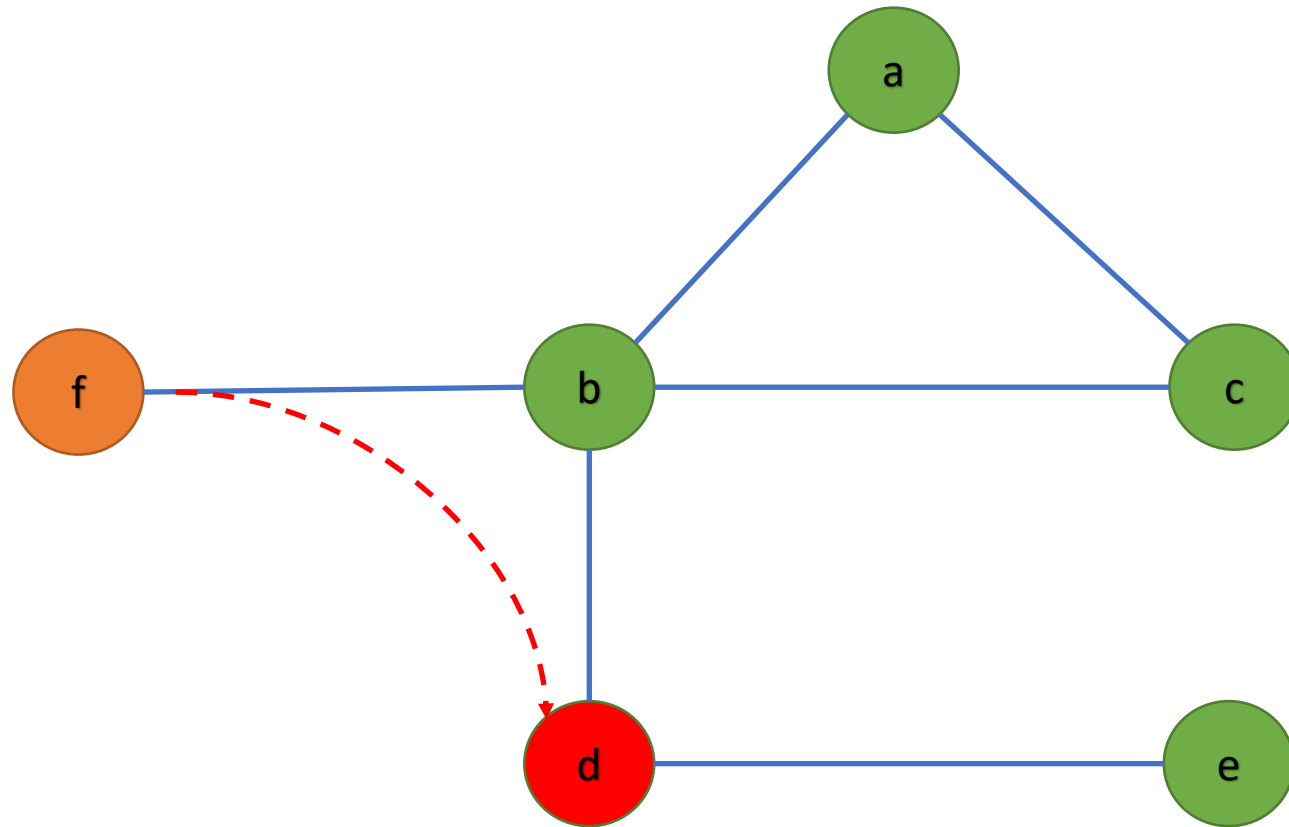
Random Walk



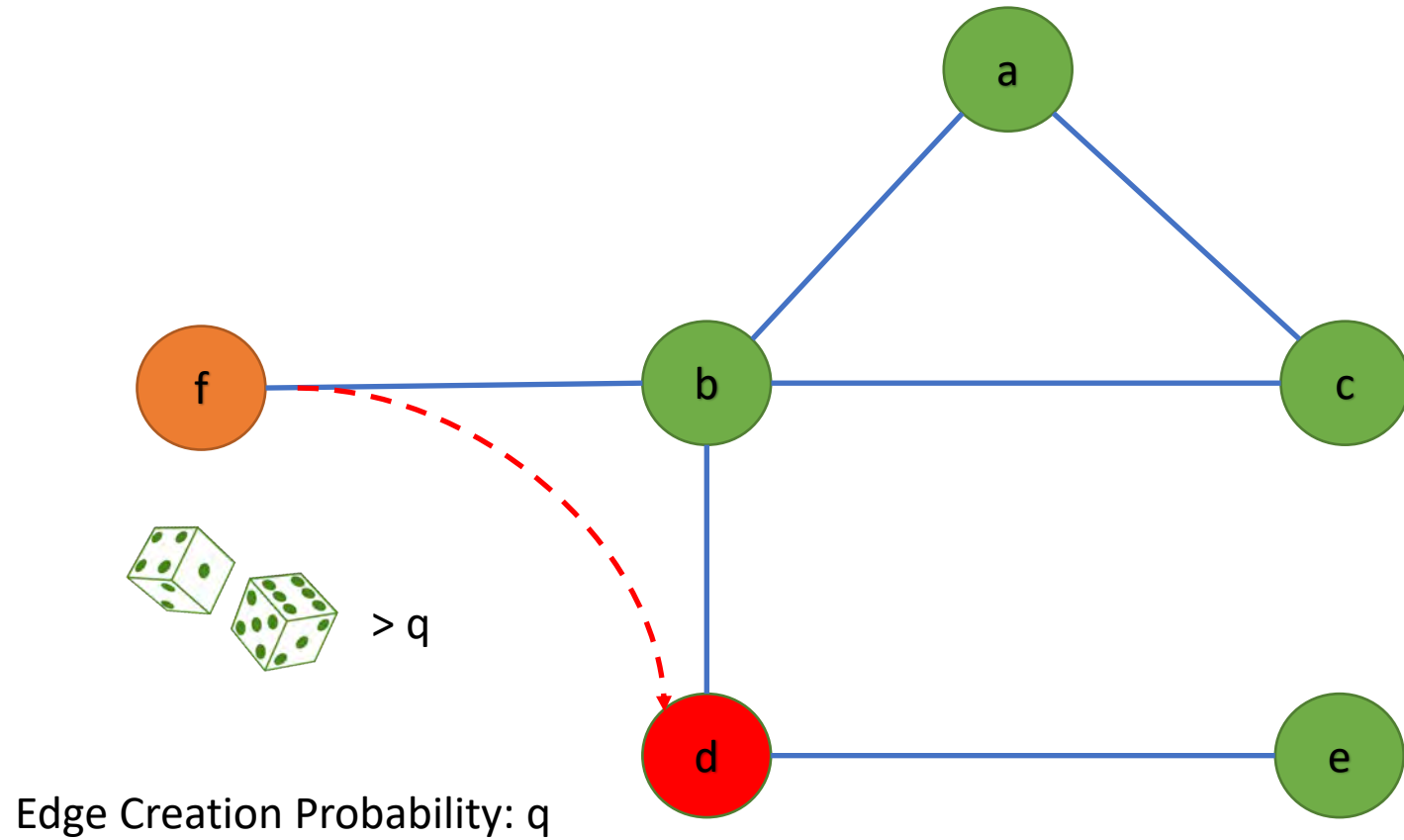
Random Walk



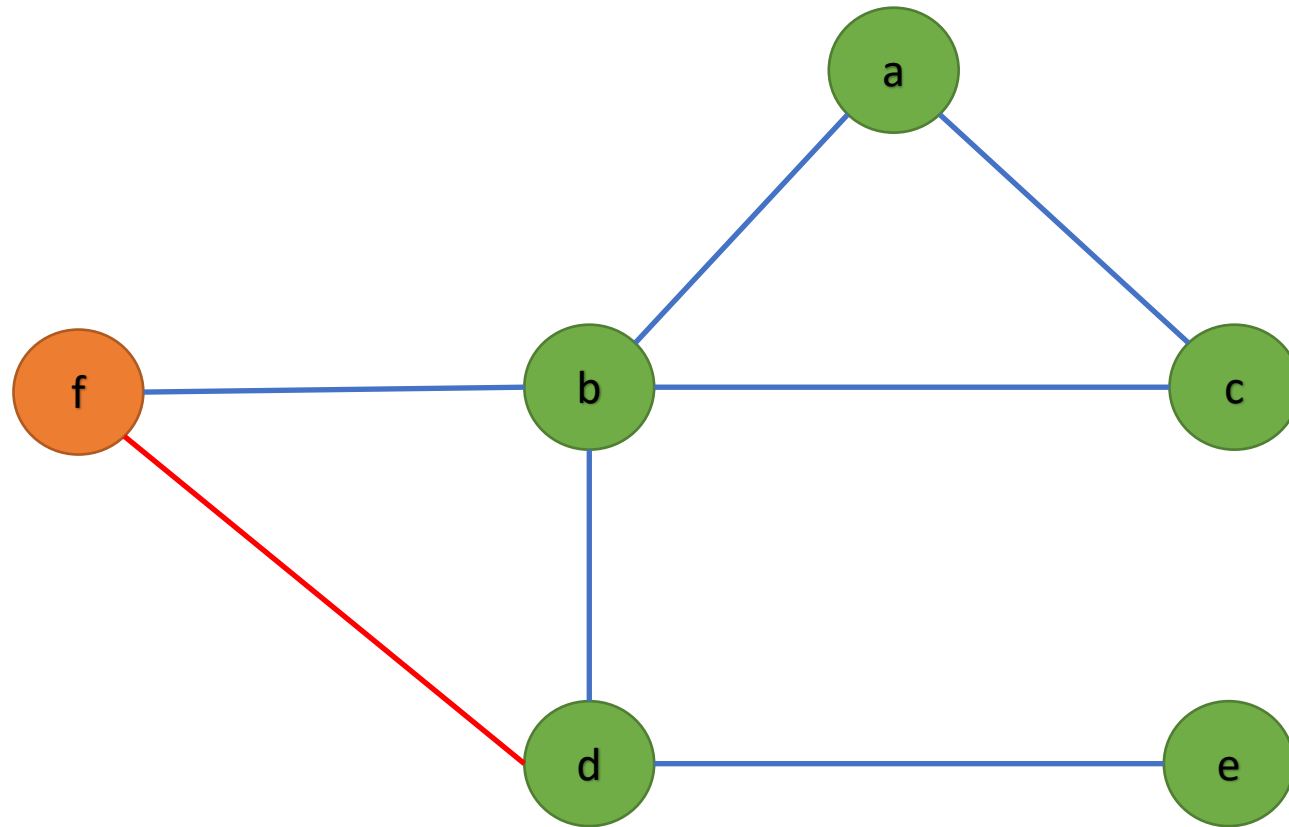
Random Walk



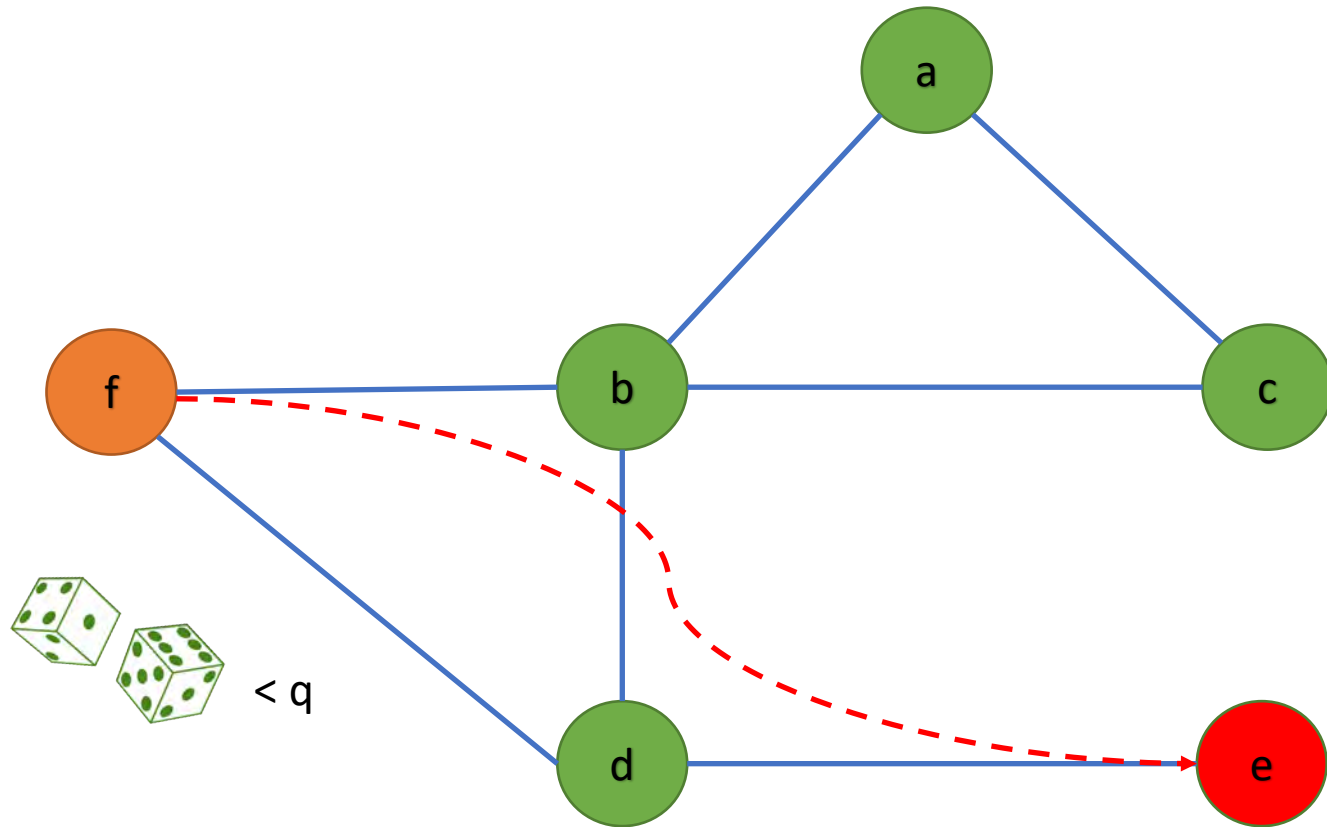
Random Walk



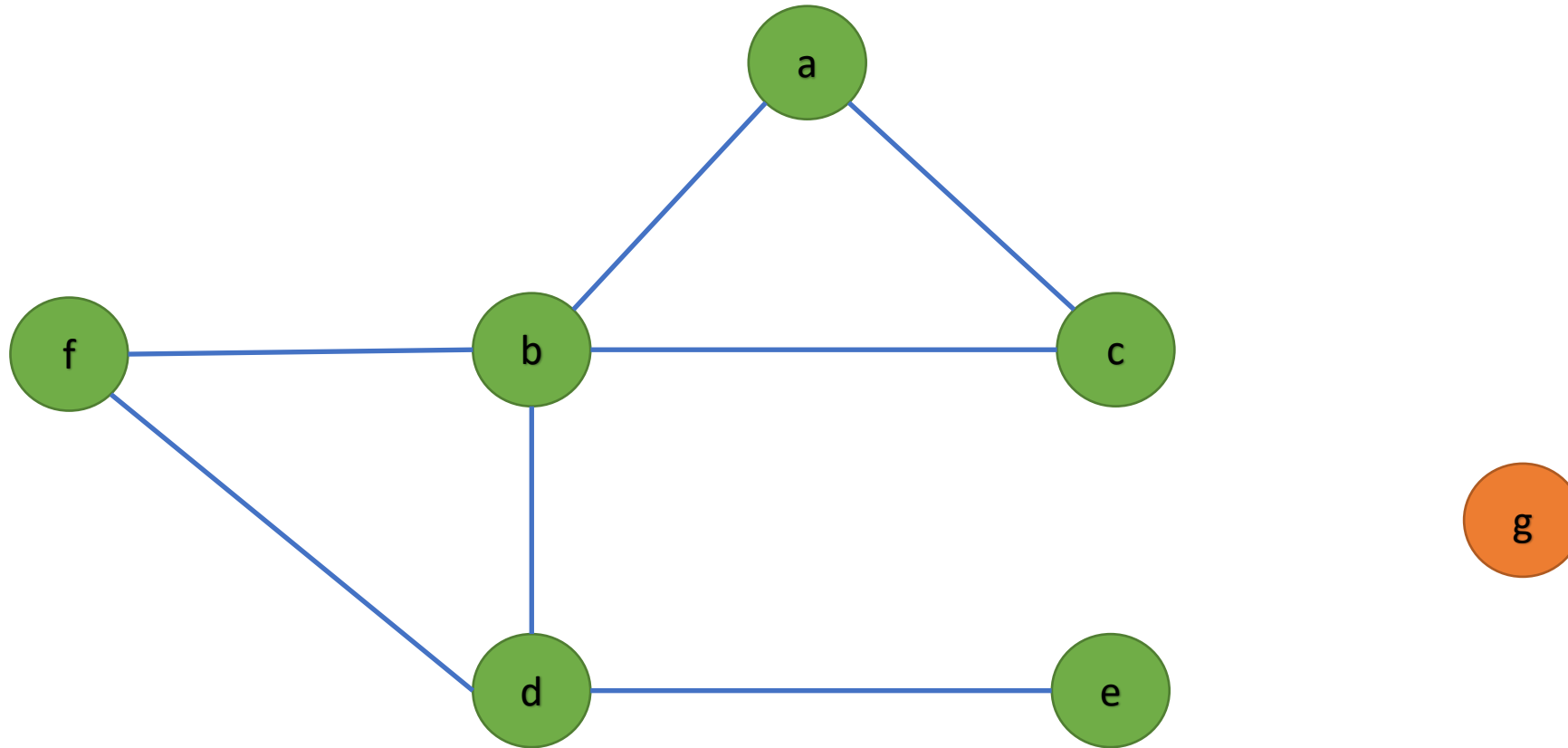
Random Walk



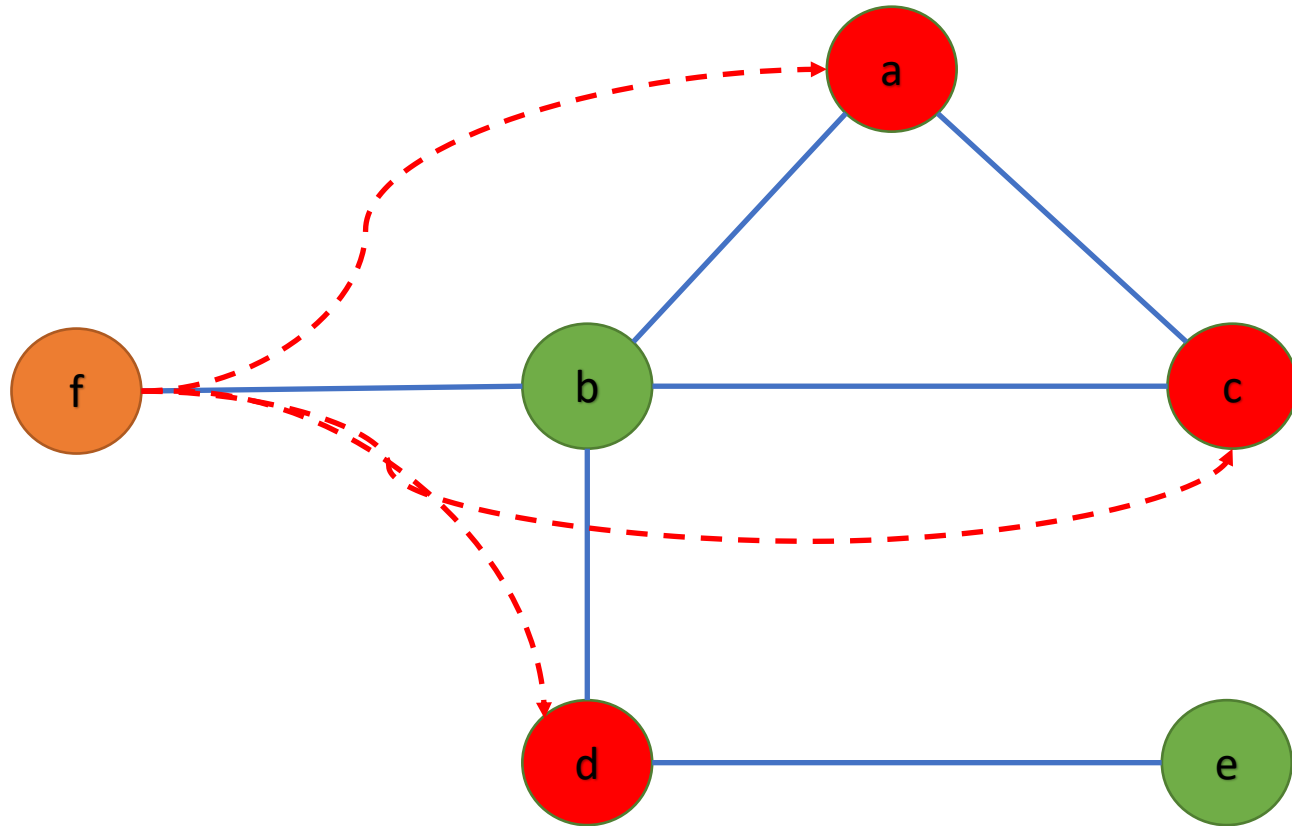
Random Walk



Random Walk

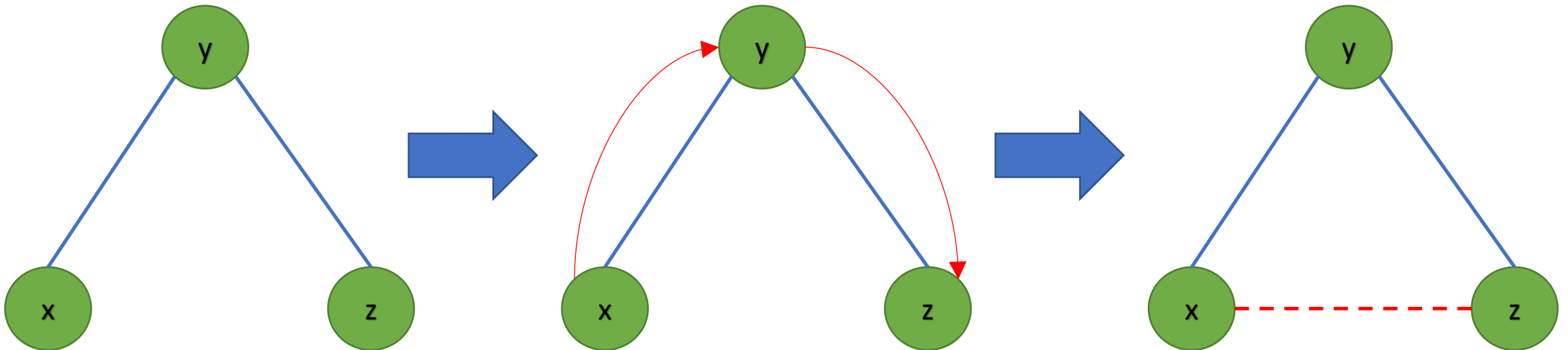


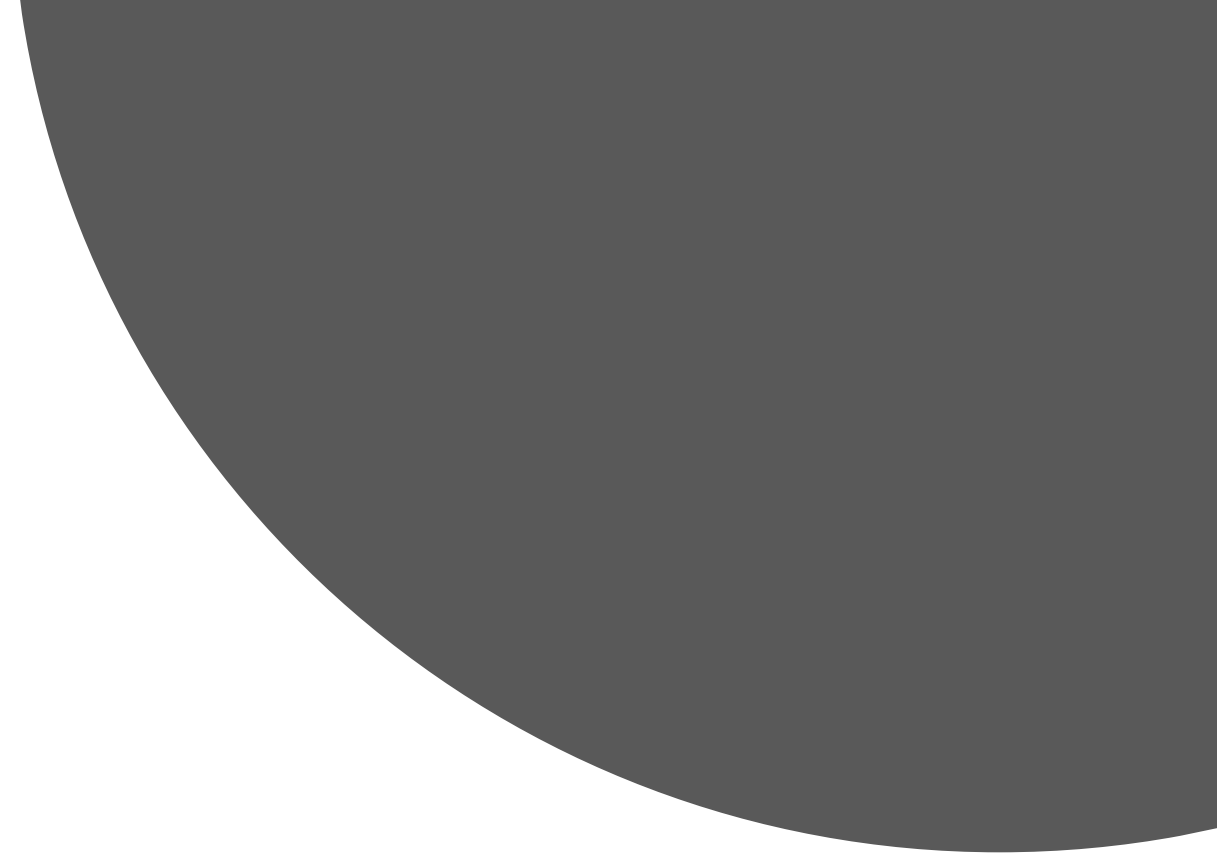
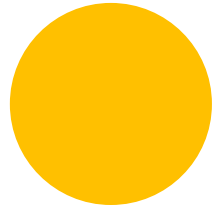
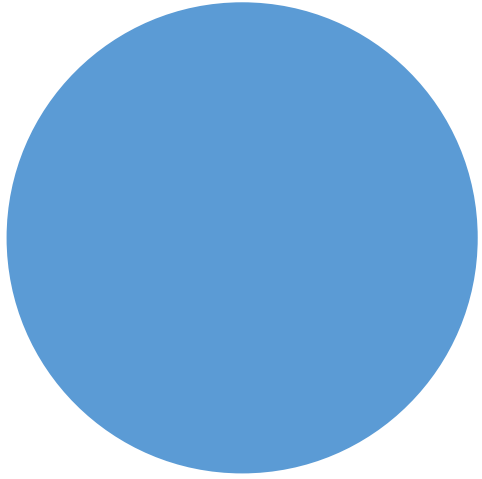
Recursive Search(RS)



Closing Triangles

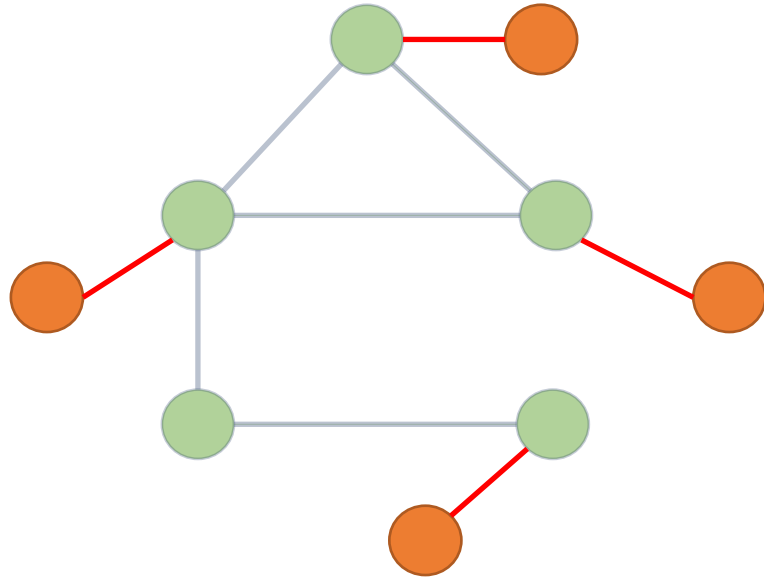
Basic block of the RW and RS can be looked as closing possible triangles





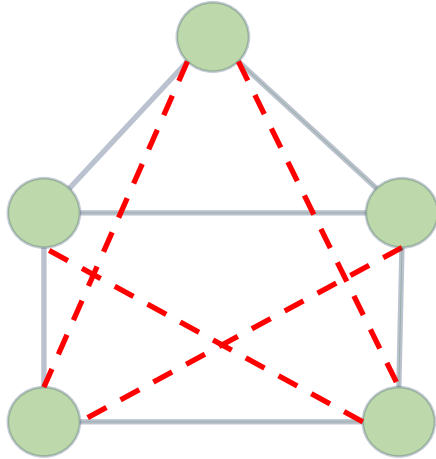
Attribute-Based Preferential Attachment (ABA)

Node Creation



- **RW and RS:** Only one new vertex will be added to the graph at each iteration
- **ABA:** Add multiple new vertices at each iteration (batch mode)

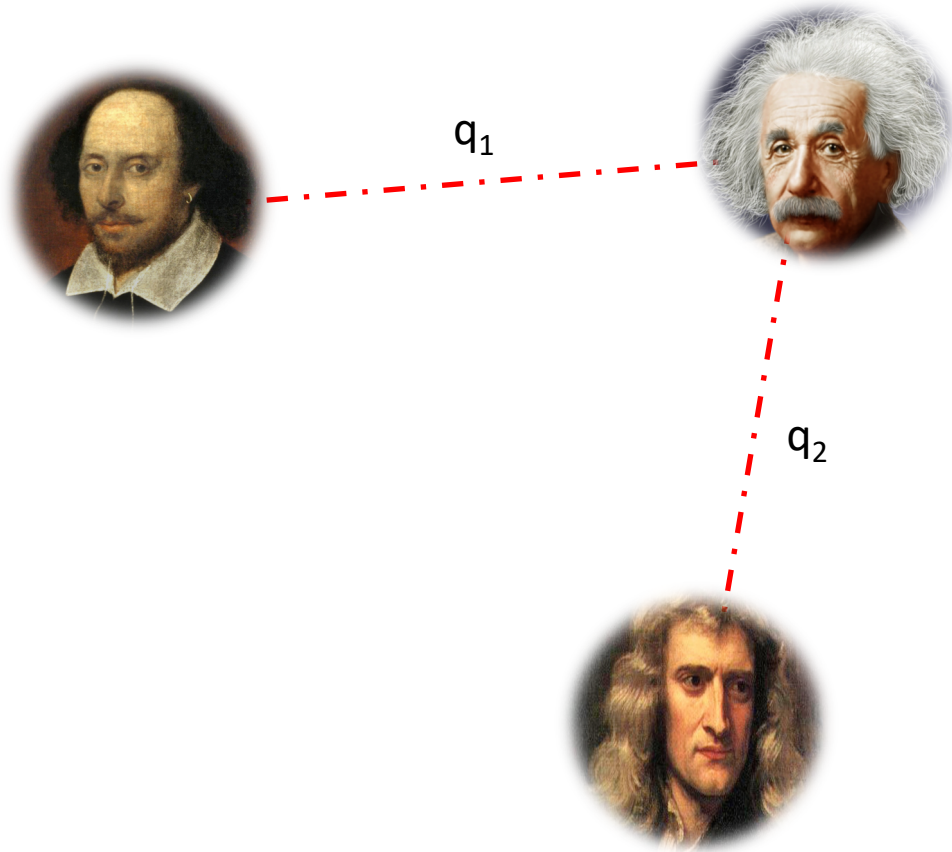
Edge Creation



RW and RS: Each vertex creates all of its edges in one single iteration

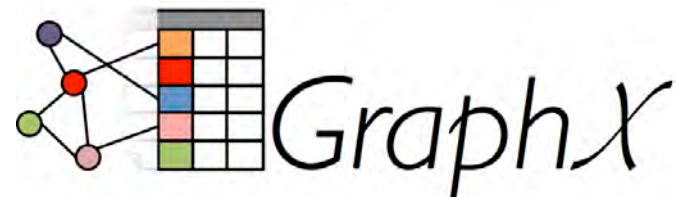
ABA: Try to close all open triangles in each iteration

Edge Creation Probability

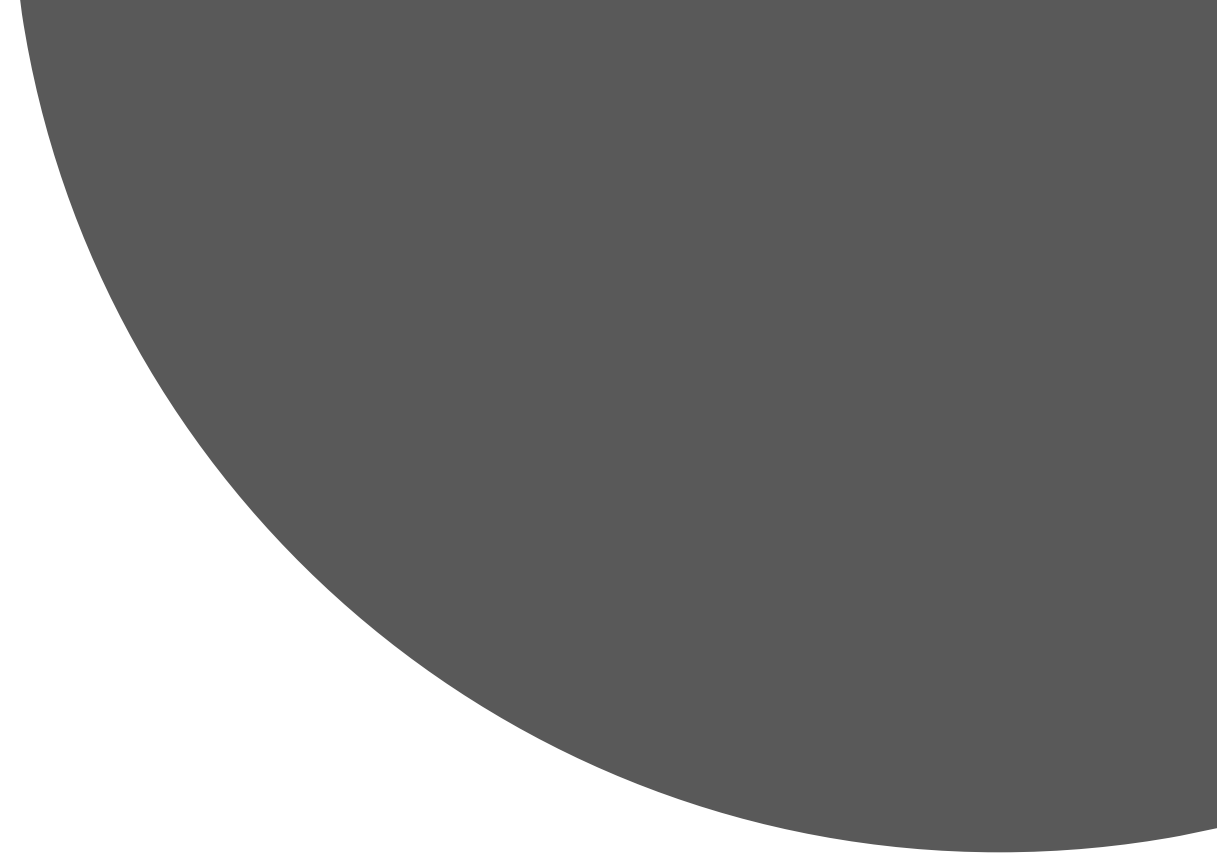
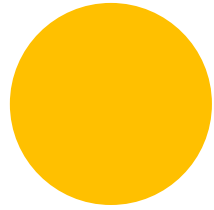
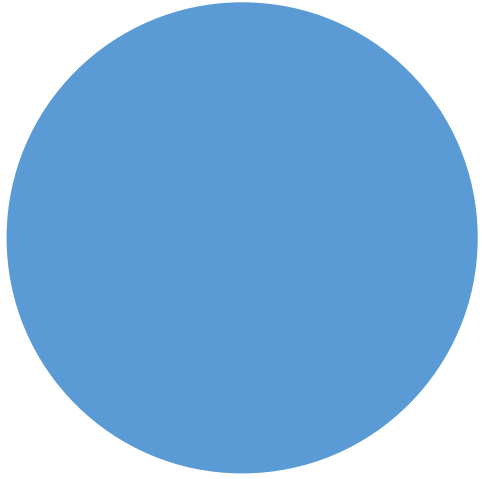


RW and RS: New edges created with a fixed probability q

ABA: Probabilities depends on pairwise node similarity



Implementation



Graph Generation

Generation Process

Node Addition

Edge Creation

Generation Process

Node Addition

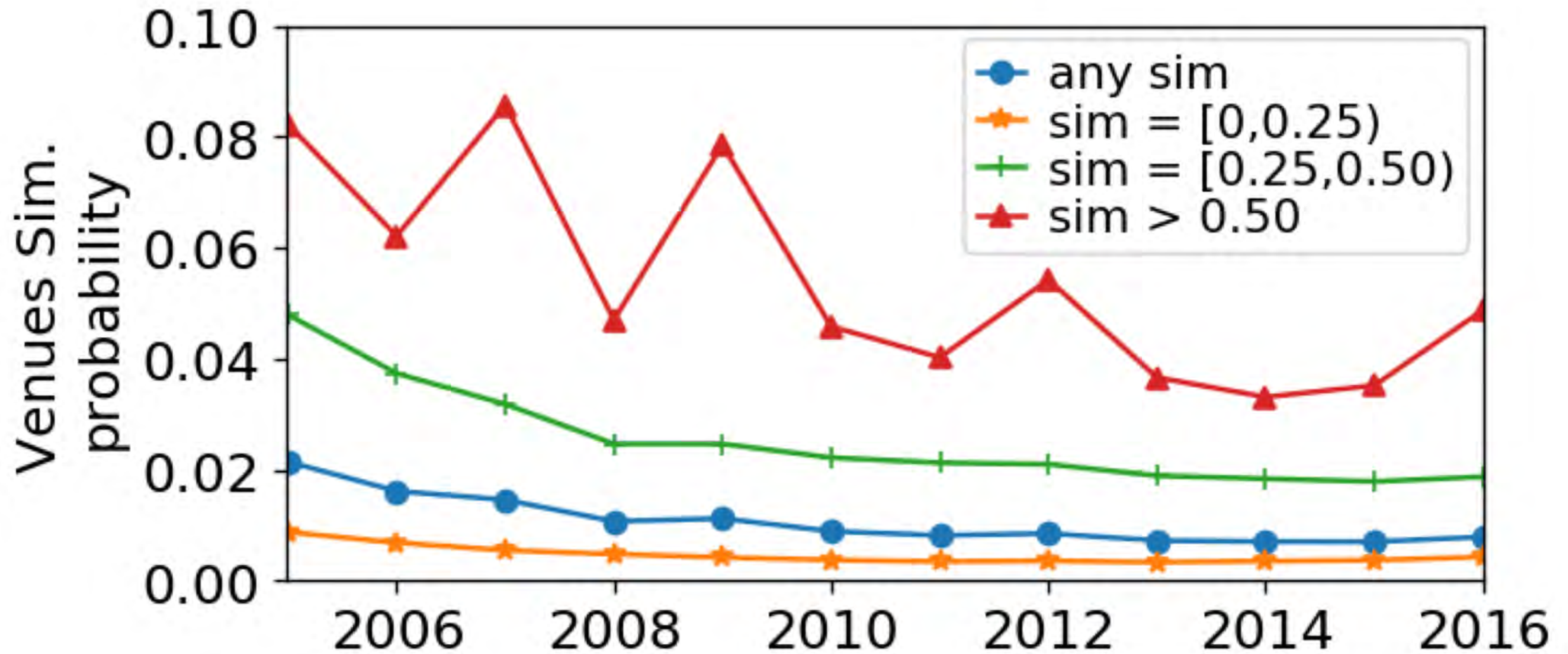
Edge Creation

Case Study: DBLP collaboration graph

- Extracted a growth-only evolving graph
 - vertices represent authors
 - evolving vertex **attribute**: venues where the author has published papers
 - edges represent co-authorship

VID	Name	Start Date	End Date	Venues
618971	Julia Stoyanovich	2005	2006	{VLDB}
618971	Julia Stoyanovich	2006	2007	{VLDB}
618971	Julia Stoyanovich	2007	2008	{VLDB, WebDB }
618971	Julia Stoyanovich	2008	2009	{VLDB, WebDB, SIGMOD , Bioinformatics }
...				

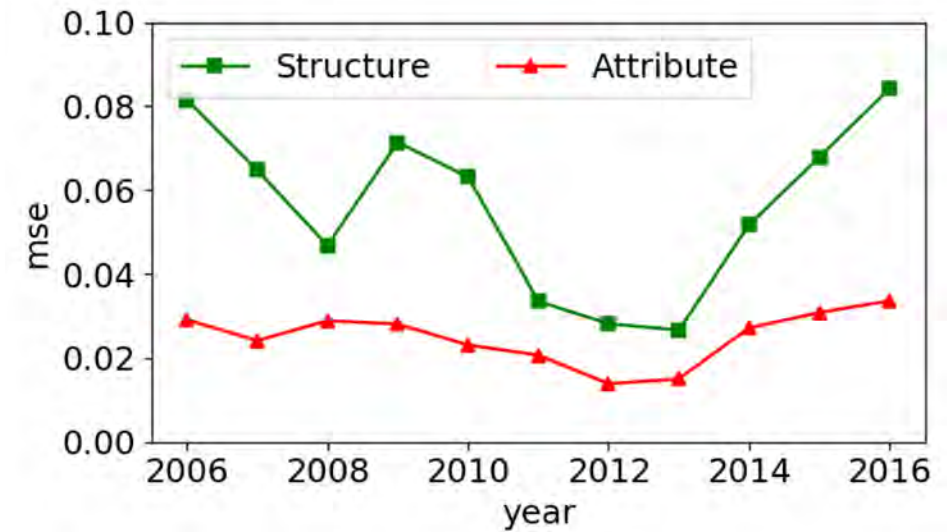
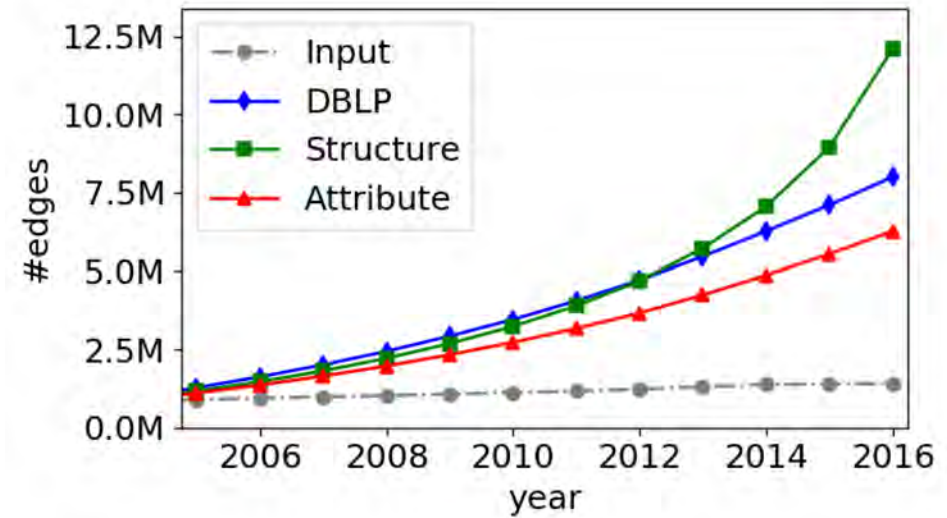




Triangle-closing probabilities in DBLP

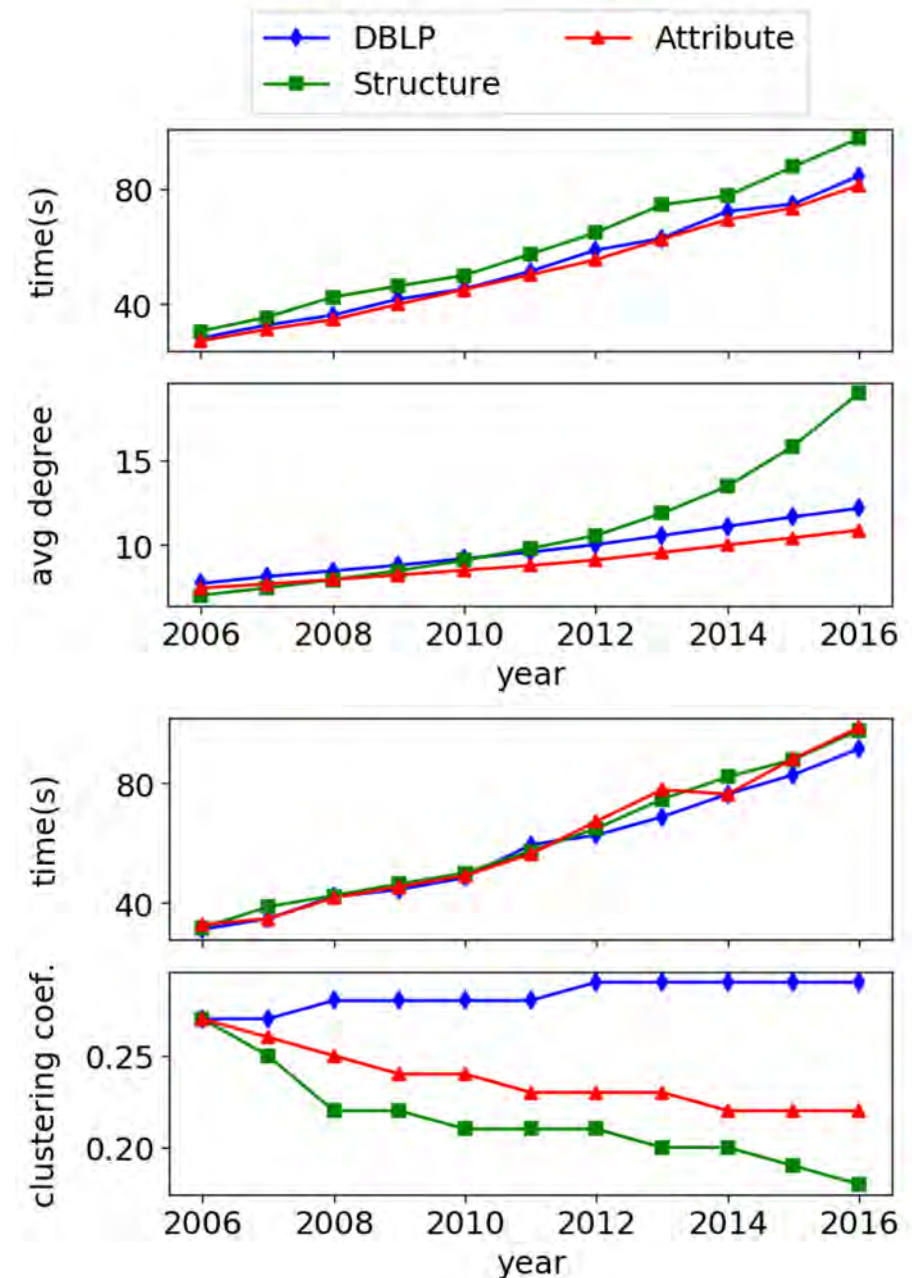
Models

- Generated using probabilities: attribute-based and structure-only



Querying the graphs

- Used Portal to run two different queries on our generated models
 - Average Vertex Degree
 - Clustering Coefficient



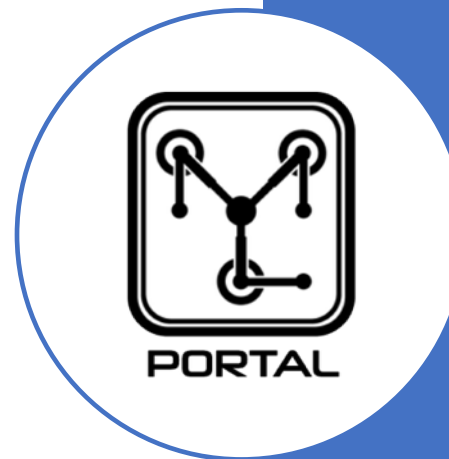
Conclusion and Future work

- In this paper we presented a data generator for evolving property graphs
- Previous models do not use attributes in their generation process
- We showed that attributes play important rule in edge creation probability
- In future works:
 - automatic methods for attaching new nodes to the graph
 - modeling attribute value evolution
 - defining a workload of evolving graph query and analysis primitives.



Thank You

<https://portaldb.github.io>



CAREER: Querying Evolving Graphs