CSCU9YQ - NoSQL Databases
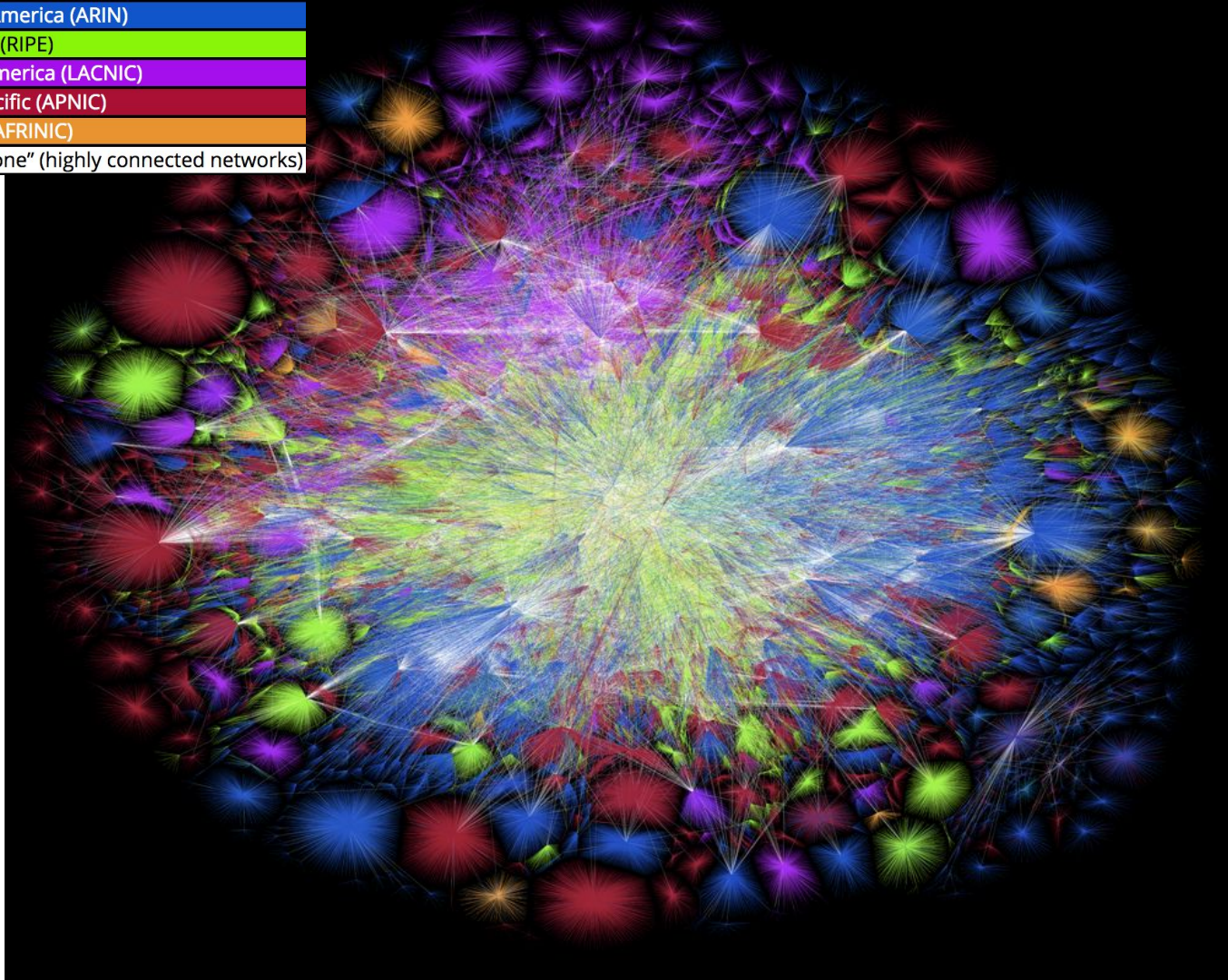
Lecture 10a: Intro to Complex Networks

Prof Gabriela Ochoa
http://www.cs.stir.ac.uk/~goc/
University of Stirling

North America (ARIN)
Europe (RIPE)
Latin America (LACNIC)
Asia Pacific (APNIC)
Africa (AFRINIC)
"Backbone" (highly connected networks)

A visual Map of the Internet

North America (ARIN)	
Europe (RIPE)	
Latin America (LACNIC)	
Asia Pacific (APNIC)	
Africa (AFRINIC)	
"Backbone" (highly connected networks)	

- Project started in 2003 by computer scientist and artist Barrett Lyon, last visualisation 2015

- Map of routing paths: paths through which information flows from router to router all across the world.

- Nodes: IP addresses (devices such as routers)

- Lines (edges): links between nodes. The length of the lines are indicative of the delay between those two nodes.
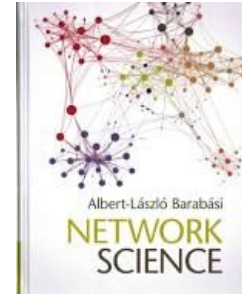
- Colour: different regions of the world

http://www.opte.org/the-internet/
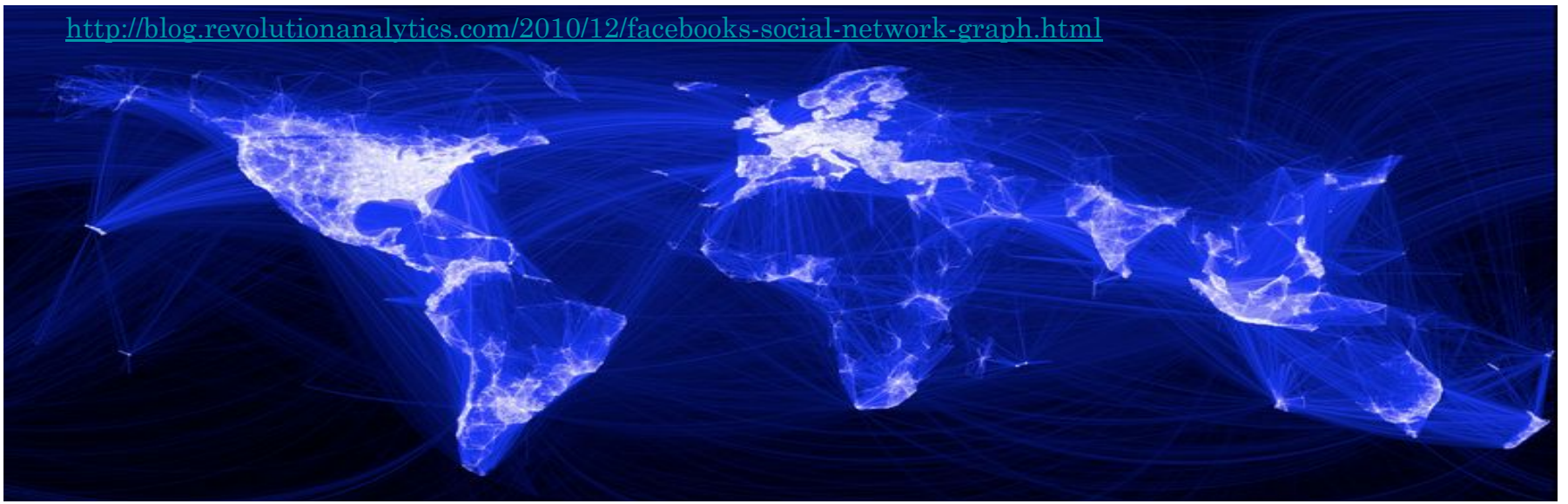
Facebook's Social Network Graph, By Paul Butler

- 10 million friend pairs from Facebook's Hadoop-based database
- Nodes: cities
- Edges: connect two cities if there is at least one friend-pair. Edge weight is proportional to the number of friends between them. Edge colours depend on their weight from black to blue to white.
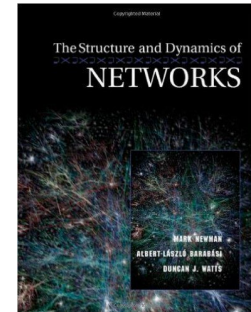
Resources

- Books
  - Network Science Book by Laszlo Barabasi et al.
    http://barabasi.com/networksciencebook/)
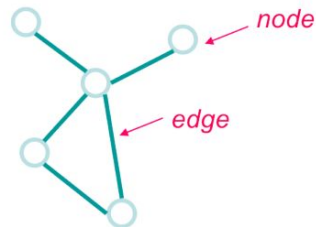 - Networks: An Introduction, M. E. J. Newman, Oxford University Press, Oxford (2010)
- Articles
  - Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2):167–256
  - Newman, M. E. (2001) The structure of scientific collaboration networks.

What is a Network?

- A collection *points* that are connected with *lines*

- A Graph is an ordered pair $G = (V, E)$, where V set of vertices, E set of edges (2-element subsets of V)

- Graph: mathematical abstraction,

- Network: real world instantiation

| Points	Lines	Discipline
vertices	edges, arcs	Math (Graph Theory)
nodes	links	Computer Science
sites	bonds	Physics
actors	ties, relations	Sociology

Terminology according to the discipline
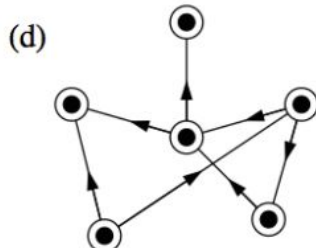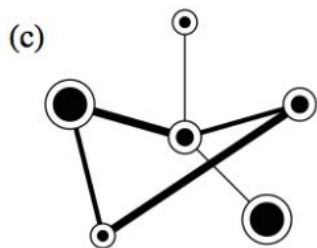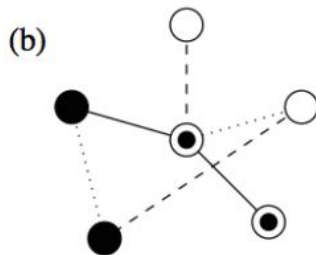
What is a Complex Network?

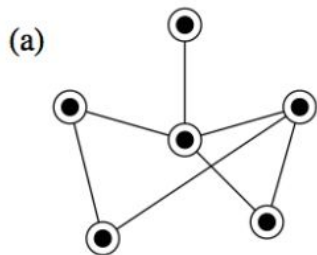- Have nontrivial topological features, i.e. its structure is irregular as opposed to regular/simple.

- Are not random either

- Can evolve over time

Regular Graphs

random network
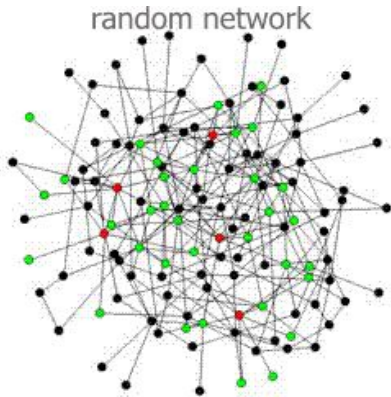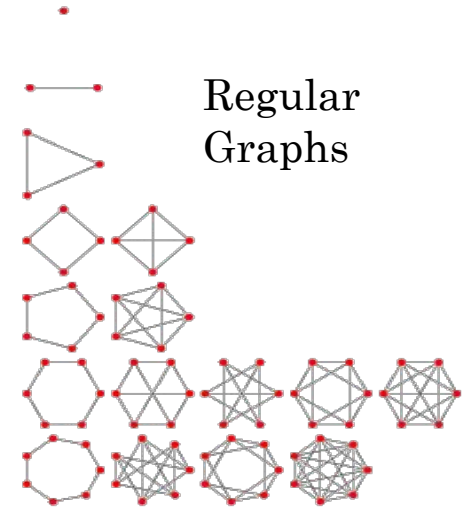
Complex Network:
Part of WWW

7
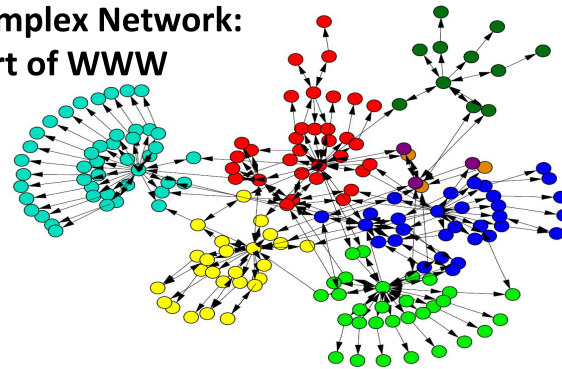
Types of networks

From (Newman, 2003)
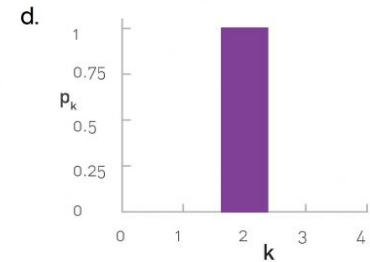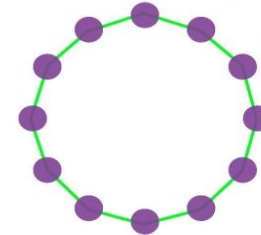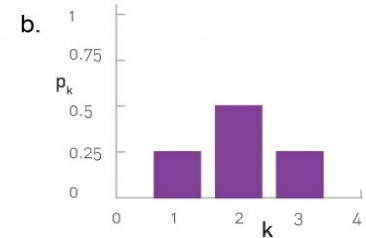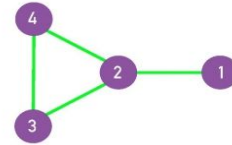
(a)   un-weighted, undirected

(b)   discrete vertex and edge types, undirected

(c)   varying vertex and edge weights, undirected

(d)   Directed (also called arcs)

General concepts: Degree and Degree Distribution

- Degree *δ(i)* of vertex *i*: number of edges incident on *i*

- Degree distribution: probability distribution of these degrees over the whole network.

General concepts: Diameter and shortest path
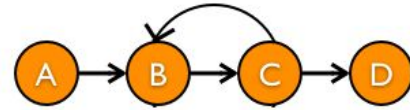
- **Shortest path**: the path between two nodes that visits the fewest intermediate nodes (path with less edges)
  - In the graph above, A->B->C->D is shorter than A->B->C->B->D (disallowing loops)
- Let $d(v_i, v_j)$ be the shortest-path distance between nodes i and j

Diameter: length of the longest shortest path between two vertices of the graph

$$D = \max d(v_i, v_j)$$

Average shortest path distance

$$l_G = \frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} d(v_i, v_j)$$

What is a social network?

- A **social network** is a collection of people, each of whom is acquainted with some subset of the others

- Represented as a set of points (or **vertices**) denoting people, joined in pairs by lines (or **edges**) denoting acquaintance.

- One could, in principle, construct the social network for a company or firm, for a school or university, or for any other community up to and including the entire world.

Agent-based Models

Mathematical Ecology

Statistical Physics

Structure of RNA

Newman M E J PNAS 2004;101:5200-5205

Collaborations among scientists at a private research institution.

Nodes in the network represent scientists, and a line between two of them indicates they coauthored a paper during the period of study.

This network seems to divide into a number of subcommunities, as indicated by the shapes and colours of the nodes

These subcommunities correspond roughly to topics of research,

12

A Larger
Co-authorship
Network
by Gabriela Ochoa and
Nadarajen Veerapen

- Genetic and Evolutionary
  Computation Conference
  (GECCO)
- Data from 2005 - 2017

Why to study social networks?

- Inherent interest in the patterns of human interaction

- Their structure has important implications for the spread of information and disease.

- For example, the average no. of acquaintances individuals have (average *degree*) might substantially influence the propagation of a rumour, a fashion, a joke, or this year's flu.

- Understanding influence and public opinion formation

Network Science & Data Science

Interdisciplinary Nature
- Network science offers a language through which different disciplines can seamlessly interact with each other.

Data driven Nature
- What distinguishes network science from graph theory is its empirical nature, i.e. its focus on data and utility. It offers insights about a system's structure or evolution.

Computational Nature
- Given the size of many of the networks we explore, and the exceptional amount of data behind them, network science offers computational challenges.

Computational Tools

Graph/Network Database Management Systems

- DSE Graph
- InfiniteGraph
- Neo4j
- OrientDB
- TitanDB

Vertex 1	Edge 1	Vertex 2
City: San Francisco Pop: 812,826	Road: I-101 Miles: 47.8	City: San Jose Pop: 967,487

Network Analysis and Visualisation Tools

- NetworkX library for Python
- igraph library for both Python and R
- Gephi visualisation and analysis software

16