Modern Discrete Probability

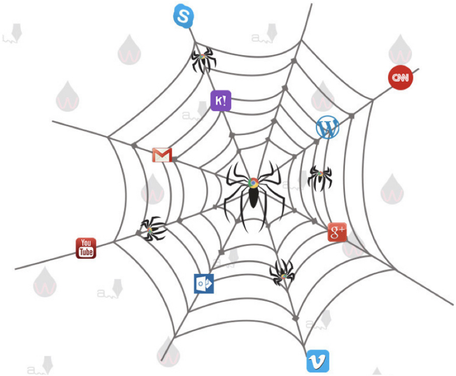
I - Introduction

Stochastic processes on graphs: models and questions

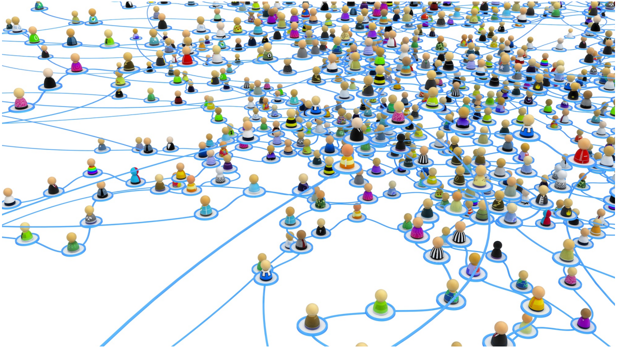
Sébastien Roch
UW–Madison
Mathematics

August 31, 2020

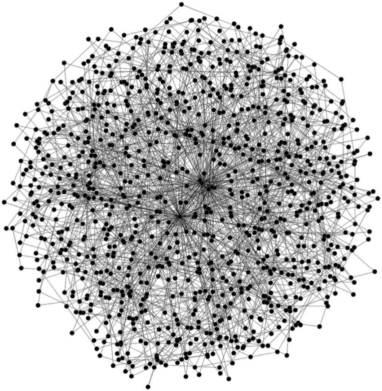
Exploring graphs



Processes on graphs



Modeling complex graphs



1. Graph terminology
2. Basic examples of stochastic processes on graphs

Graphs

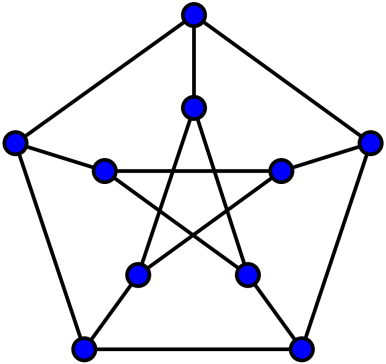
Definition

An *(undirected) graph* is a pair $G = (V, E)$ where V is the set of *vertices* and

$$E \subseteq \{\{u, v\} : u, v \in V\},$$

is the set of *edges*.

An example: the Petersen graph



Basic definitions

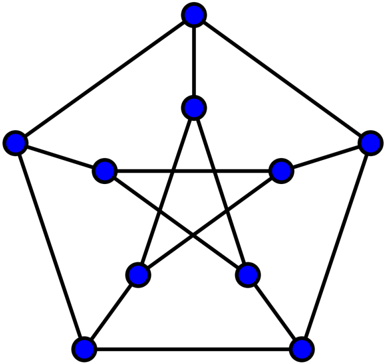
Definition (Neighborhood)

Two vertices $u, v \in V$ are *adjacent*, denoted by $u \sim v$, if $\{u, v\} \in E$. The set of adjacent vertices of v, denoted by $N(v)$, is called the *neighborhood* of v and its size, i.e. $\delta(v) := |N(v)|$, is the *degree* of v. A vertex v with $\delta(v) = 0$ is called *isolated*.

Example

All vertices in the Petersen graph have degree 3. In particular there is no isolated vertex.

An example: the Petersen graph



Paths and connectivity

Definition (Paths)

A *path* in G is a sequence of vertices $x_0 \sim x_1 \sim \cdots \sim x_k$. The number of edges, k, is called the *length* of the path. If $x_0 = x_k$, we call it a *cycle*. We write $u \leftrightarrow v$ if there is a path between u and v. The equivalence classes of $\leftrightarrow$ are called *connected components*. The length of the shortest path between two vertices u, v is their *graph distance*, denoted $d_G(u, v)$.

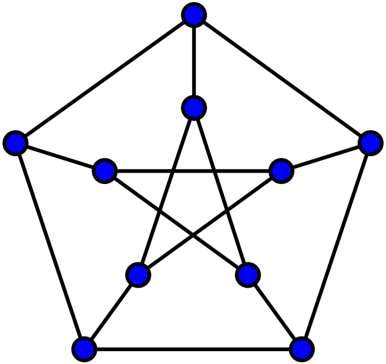
Definition (Connectivity)

A graph is *connected* if any two vertices are linked by a path, i.e., if $u \leftrightarrow v$ for all $u, v \in V$.

Example

The Petersen graph is connected.

An example: the Petersen graph



Adjacency matrix

Definition

Let $G = (V, E)$ be a graph with $n = |V|$. The *adjacency matrix* A of G is the $n \times n$ matrix defined as $A_{xy} = 1$ if $\{x, y\} \in E$ and 0 otherwise.

Example

The adjacency matrix of a *triangle* (i.e. 3 vertices with all edges) is

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Examples of finite graphs

- K_n: clique with n vertices, i.e., graph with all edges present
- C_n: cycle with n non-repeated vertices
- $\mathbb{H}^n$: n-dimensional hypercube, i.e., $V = \{0, 1\}^n$ and $u \sim v$ if u and v differ at one coordinate

Random walk on a graph

Definition

Let $G = (V, E)$ be a countable graph where every vertex has finite degree. Let $c : E \to \mathbb{R}_+$ be a positive edge weight function on G. We call $\mathcal{N} = (G, c)$ a *network*. Random walk on $\mathcal{N}$ is the process on V, started at an arbitrary vertex, which at each time picks a neighbor of the current state proportionally to the weight of the corresponding edge.

Questions:

- How often does the walk return to its starting point?
- How long does it take to visit all vertices once or a particular subset of vertices for the first time?
- How fast does it approach equilibrium?

Undirected graphical models I

Definition

Let S be a finite set and let $G = (V, E)$ be a finite graph. Denote by $\mathcal{K}$ the set of all cliques of G. A positive probability measure μ on $\mathcal{X} := S^V$ is called a *Gibbs random field* if there exist *clique potentials* $\phi_K : S^K \to \mathbb{R}$, $K \in \mathcal{K}$, such that

$$\mu(x) = \frac{1}{\mathcal{Z}} \exp\left(\sum_{K \in \mathcal{K}} \phi_K(x_K)\right),$$

where x_K is x restricted to the vertices of K and $\mathcal{Z}$ is a normalizing constant.

Undirected graphical models II

Example

For $\beta > 0$, the *ferromagnetic Ising model* with inverse temperature β is the Gibbs random field with $S := \{-1, +1\}$, $\phi_{\{i,j\}}(\sigma_{\{i,j\}}) = \beta\sigma_i\sigma_j$ and $\phi_K \equiv 0$ if $|K| \neq 2$. The function $\mathcal{H}(\sigma) := -\sum_{\{i,j\}\in E} \sigma_i\sigma_j$ is known as the *Hamiltonian*. The normalizing constant $\mathcal{Z} := \mathcal{Z}(\beta)$ is called the *partition function*. The states $(\sigma_i)_{i\in V}$ are referred to as *spins*.

Questions:

- How fast is correlation decaying?
- How to sample efficiently?
- How to reconstruct the graph from samples?

Erdös-Rényi random graph

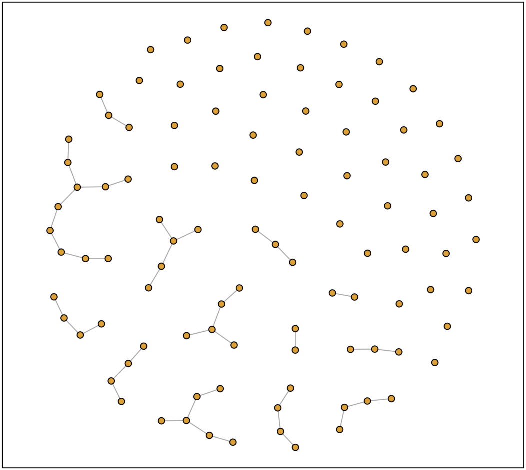
Definition

Let $V = [n]$ and $p \in [0,1]$. The *Erdös-Rényi graph* $G = (V, E)$ on n vertices with density p is defined as follows: for each pair $x \neq y$ in V, the edge $\{x, y\}$ is in E with probability p independently of all other edges. We write $G \sim \mathbb{G}_{n,p}$ and we denote the corresponding measure by $\mathbb{P}_{n,p}$.

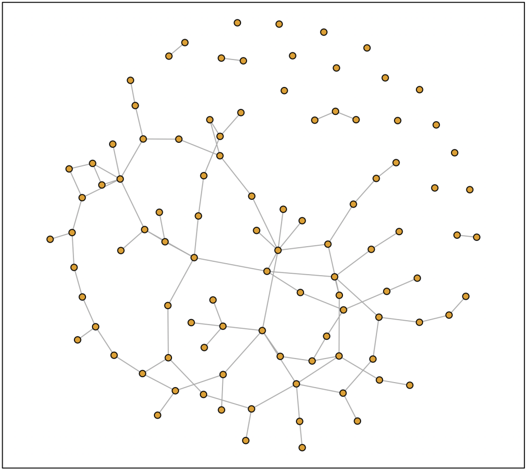
Questions:

- What is the probability of observing a triangle?
- Is G connected?
- What is the typical chromatic number (i.e., the smallest number of colors needed to color the vertices so that no two adjacent vertices share the same color)?

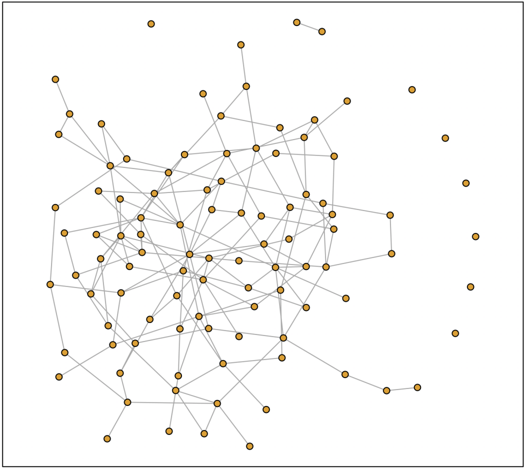
Erdös-Rényi with $n = 100$ and $p_n = 1/100$



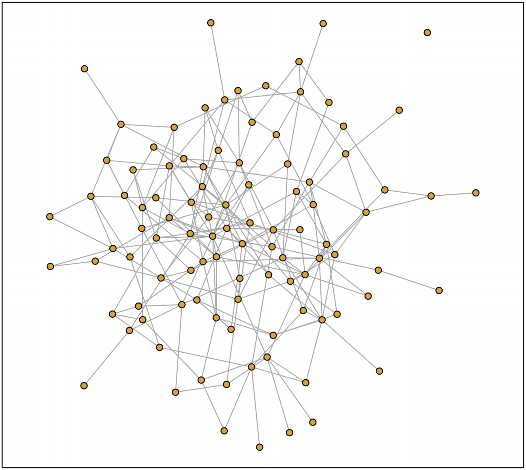
Erdös-Rényi with $n = 100$ and $p_n = 2/100$



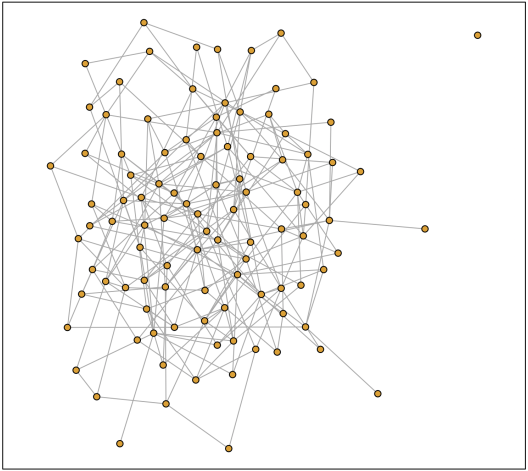
Erdös-Rényi with $n = 100$ and $p_n = 3/100$



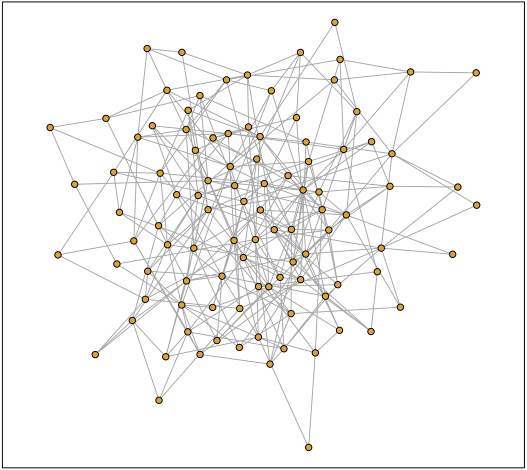
Erdös-Rényi with $n = 100$ and $p_n = 4/100$



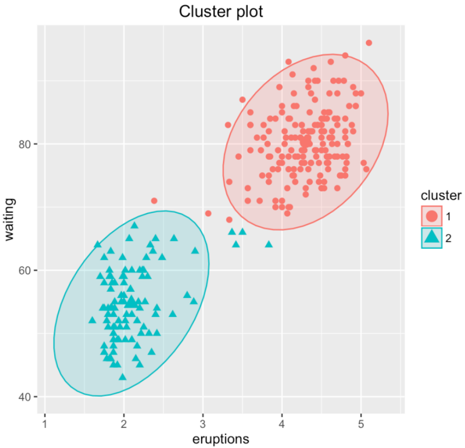
Erdös-Rényi with $n = 100$ and $p_n = 5/100$



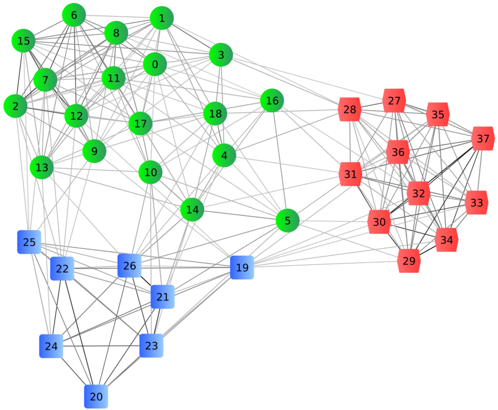
Erdös-Rényi with $n = 100$ and $p_n = 6/100$



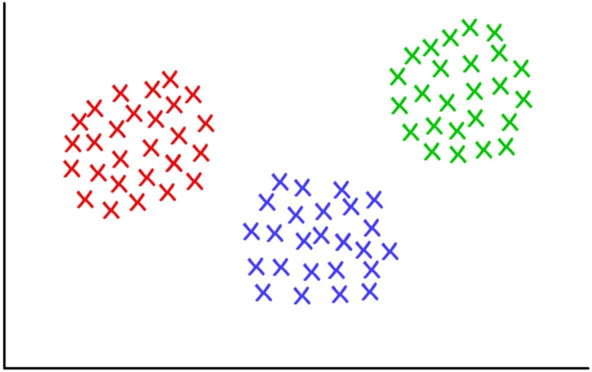
Clustering in Euclidean space



Clustering in graphs



Reducing the second problem to the first one



Go deeper

More details at:

`http://www.math.wisc.edu/~roch/mdp/`