

CSC17103 – KHAI THÁC DỮ LIỆU ĐỒ THỊ

HOMEWORK 01: MẪU ĐỒ THỊ & PHÁT SINH ĐỒ THỊ

1 Bài tập

Phát sinh mạng Erdős-Rényi

Hãy trực quan hoá mạng Erdős-Rényi với $N = 50$ nút và bậc trung bình $\langle k \rangle$ lần lượt là

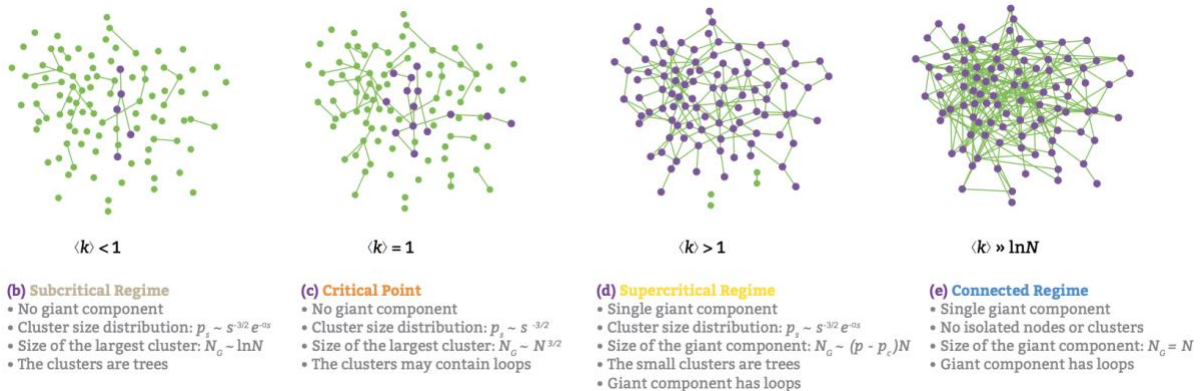
a. $\langle k \rangle = 0.5$

b. $\langle k \rangle = 4$

c. $\langle k \rangle = 8$

Lưu ý: sử dụng thư viện `networkx`. Chèn hình và code tương ứng với từng trường hợp.

Mạng Erdős-Rényi



Hình 1. Phân loại chế độ topo khi thay đổi $\langle k \rangle$

Xem xét mạng $G(N, p)$ có $N = 3000$ nút và được kết nối với nhau với xác suất $p = 10^{-3}$.

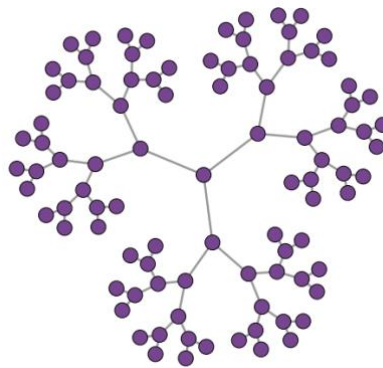
Hãy trả lời các câu hỏi dưới đây.

- Xác định số lượng liên kết kỳ vọng $\langle L \rangle$ và bậc trung bình $\langle k \rangle$ của mạng.
- Xác suất có chính xác 50 liên kết trong mạng là bao nhiêu ?
- Dựa vào hình 1, xác định xem mạng ở chế độ (regime) nào ?
- Tính xác suất p_c để mạng ở chế độ critical point.
- Tính số nút N^{cr} , bậc trung bình $\langle k^{cr} \rangle$ và khoảng cách trung bình giữa hai nút được chọn ngẫu nhiên $\langle d \rangle$ để mạng chỉ có một thành phần.
- Tìm phân bố bậc p_k của mạng này (xấp xỉ với phân bố bậc Poisson).

Cây Cayley (Cayley tree)

Cây Cayley là cây đối xứng, được xây dựng bắt đầu từ nút trung tâm bậc k . Mỗi nút ở khoảng cách d tính từ nút trung tâm có bậc k , cho đến khi chúng ta đến các nút ở khoảng cách P có bậc một và được gọi là các lá. Ví dụ, hình 2 là cây Cayley có $k = 3$ và $P = 5$.

- Tính tổng số nút trên cây sau t bước tính từ nút trung tâm.
- Tính độ phân phối bậc (degree distribution) của mạng.
- Tính đường kính d_{\max} .
- Tìm biểu thức của đường kính d_{\max} theo tổng số nút N .



Hình 2. Cây Cayley

Nghịch lý tình bạn (Friendship Paradox)

Phân phối bậc p_k là xác suất mà một nút được chọn ngẫu nhiên có k hàng xóm. Tuy nhiên, nếu chúng ta chọn ngẫu nhiên một liên kết, xác suất để một nút ở một trong các đầu của nó có bậc k là $q_k = A p_k$, trong đó A là hệ số chuẩn hóa.

- Tìm hệ số chuẩn hóa A , giả sử rằng mạng có phân bố bậc theo luật mũ với $2 < \gamma < 3$, với bậc nhỏ nhất k_{\min} và bậc lớn nhất k_{\max} .
- Chọn ngẫu nhiên một nút trong mạng có $N = 104$, $\gamma = 2.3$, $k_{\min} = 1$ và $k_{\max} = 1000$. Tính bậc trung bình của các nút lân cận.

2 Tài liệu tham khảo

[1] Network Science by Albert-László Barabási. (n.d.). BarabásiLab. <http://networksciencebook.com>

[2] Chakrabarti, D., & Faloutsos, C. (2012, October 30). Graph Mining: Laws, Tools, and Case Studies.