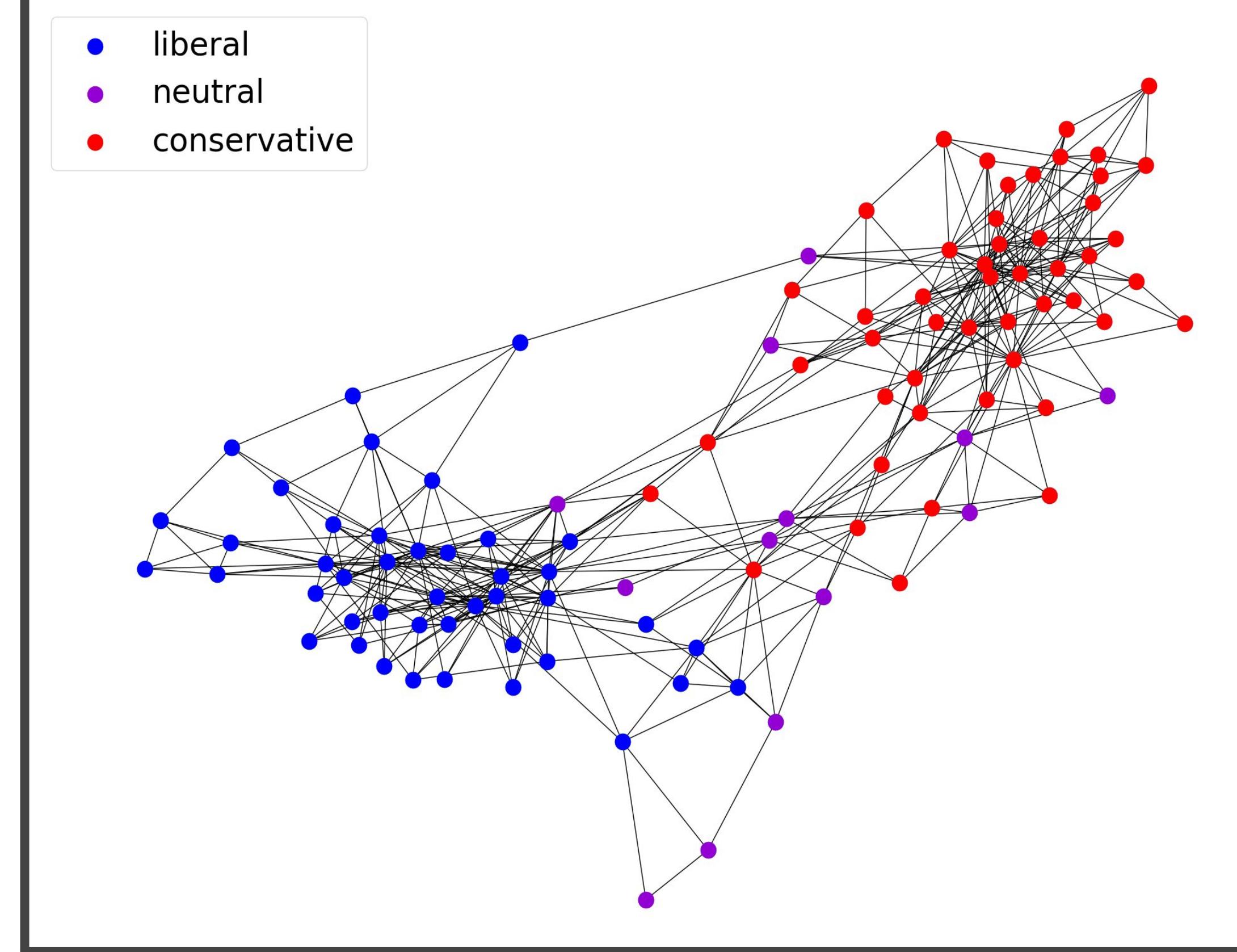


Network of Books on Politics

Data Description ♣

- 무방향 단순 네트워크
(multi-edge, self-loop X)
- [amazon.com](#)에서 살 수 있는 정치 도서를 노드로 하고, 같은 구매자에게 함께 구매된 도서 간에 엣지를 이은 네트워크
- 약 2004년에 수집됨.
- 진보(● liberal), 중도(● neutral), 보수(● conservative)의 세 가지 도서 유형으로 구분됨.

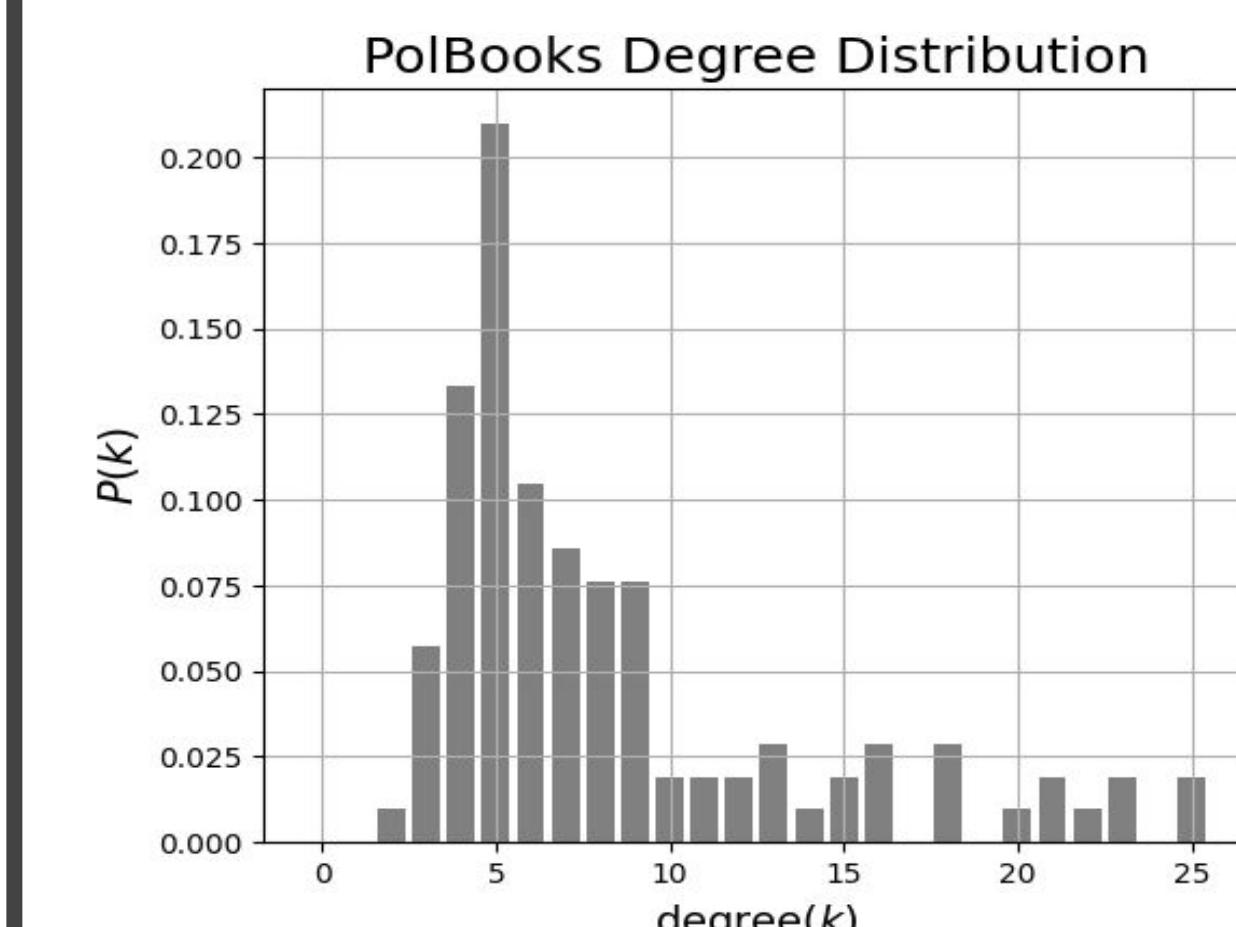
노드 유형	정치 도서
엣지 유형	공동 구매 이력
노드수	105
엣지수	441



Original Data Properties ↔

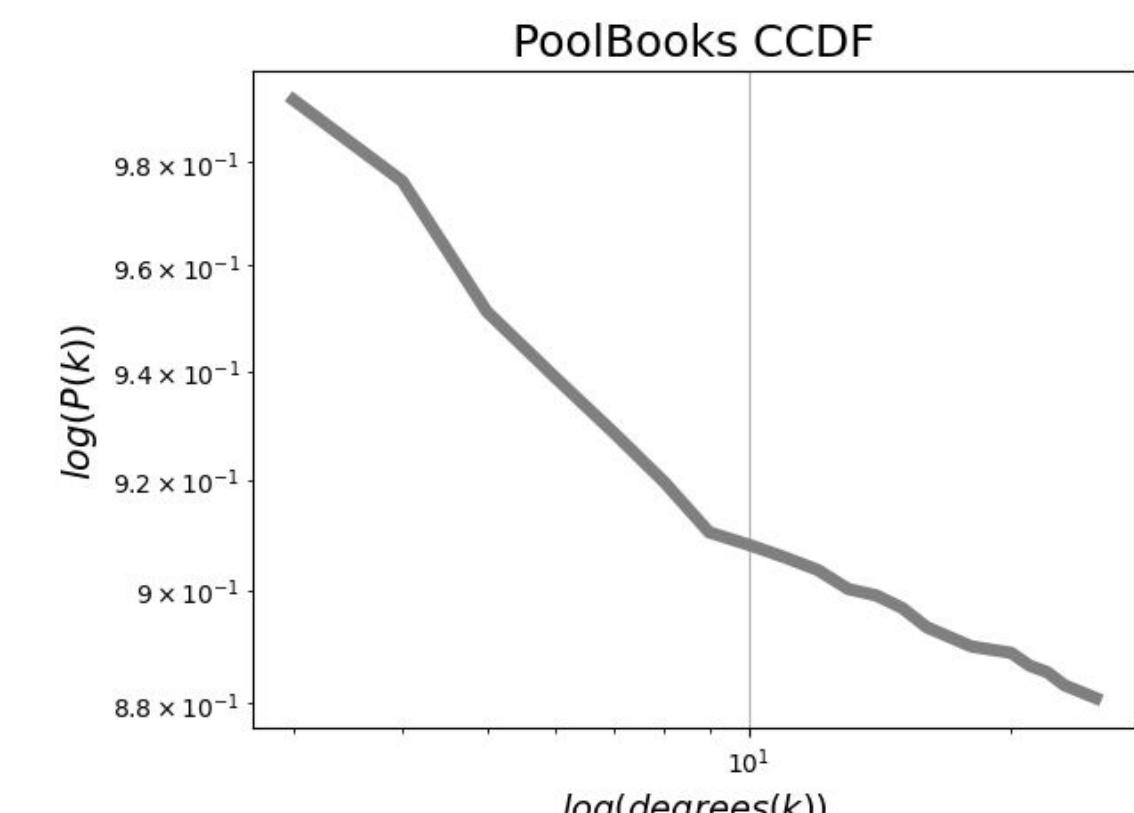
평균 차수 : 각 노드가 갖는 평균 차수(이웃수)
 $\langle k \rangle = \frac{2E}{N} = 8.4$

엣지 밀도 : 네트워크가 가질 수 있는 엣지수 대비 실제 엣지수
 $D = \frac{2E}{N(N-1)} = 0.08$
 (N : 노드 수 / E : 엣지 수)



- 밀도가 1보다 매우 작기 때문에 해당 네트워크는 조밀하게 연결되어 있지 않음.

- CCDF를 통해 두꺼운 꼬리 분포(heavy-tailed)를 확인할 수 있으므로 차수가 매우 큰 노드, 즉 허브가 존재함을 알 수 있음.



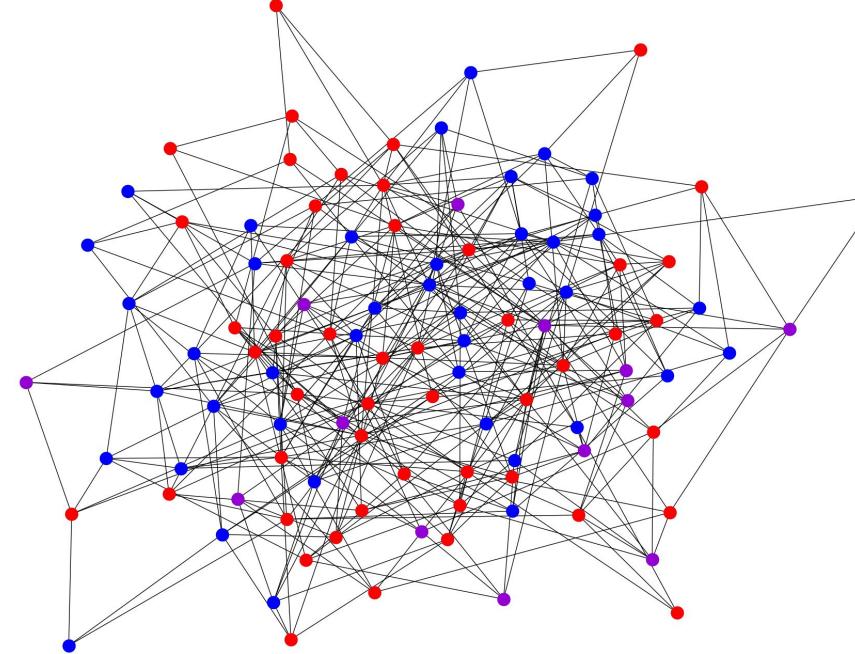
Random Models

원본 네트워크에서 특정한 특성을 유지한 무작위 네트워크를 생성하여 원본 네트워크의 특성을 분석. 세 가지 무작위 모델에서 모두 원본 네트워크의 그룹 분리 현상은 재현되지 않음.

ER model? ↔

: 노드 간 확률 p 로 엣지를 연결하여 그래프를 생성하는 모델

$$p = \frac{\langle k \rangle}{(N-1)} = 0.08$$

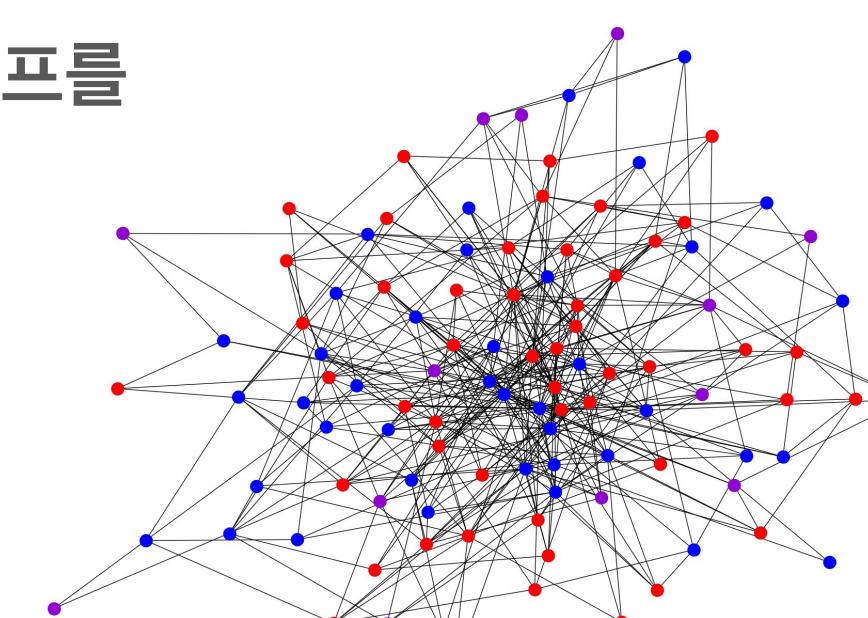


유지되는 특성
- 노드수
- 엣지 밀도

Configuration model? ♣

: 각 노드의 차수를 고정하고 엣지를 무작위 연결하여 그래프를 생성하는 모델

유지되는 특성
- 노드수
- 엣지 밀도
- 차수

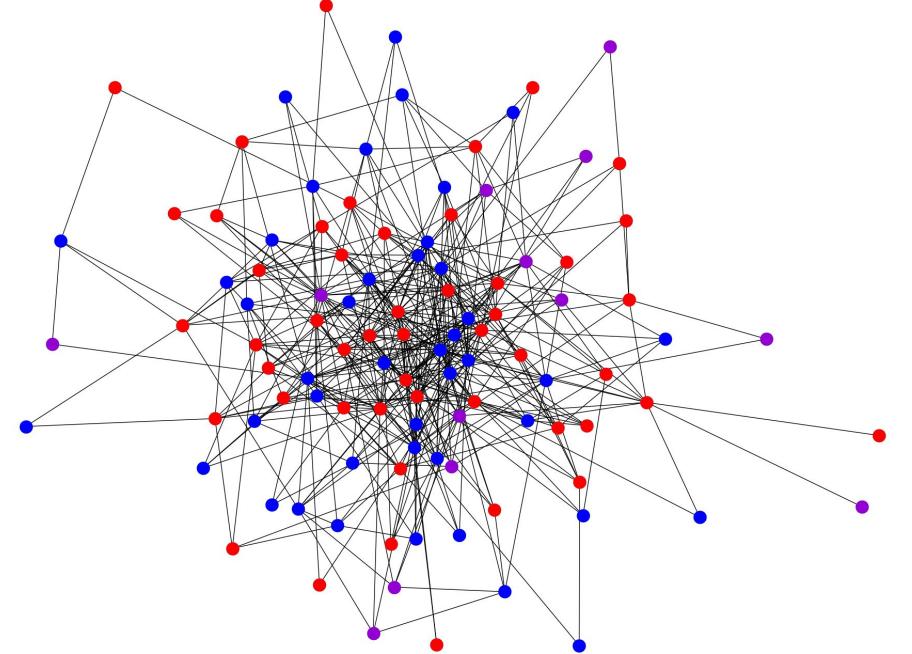


Chung-Lu model? ★

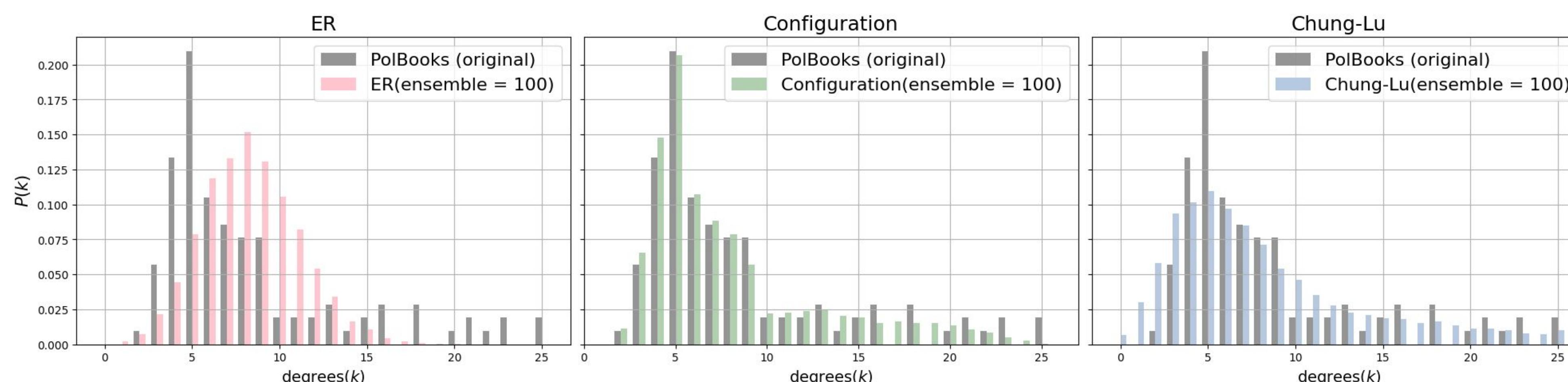
: 노드 i, j 의 차수에 따른 확률 p_{ij} 로 엣지를 연결하여 그래프를 생성하는 모델

유지되는 특성
- 노드수
- 엣지 밀도
- 차수(근사)

$$p_{ij} = \frac{k_i k_j}{\sum_i k_i}$$



Degree Distribution ♣



차수 분포(Degree Distribution)

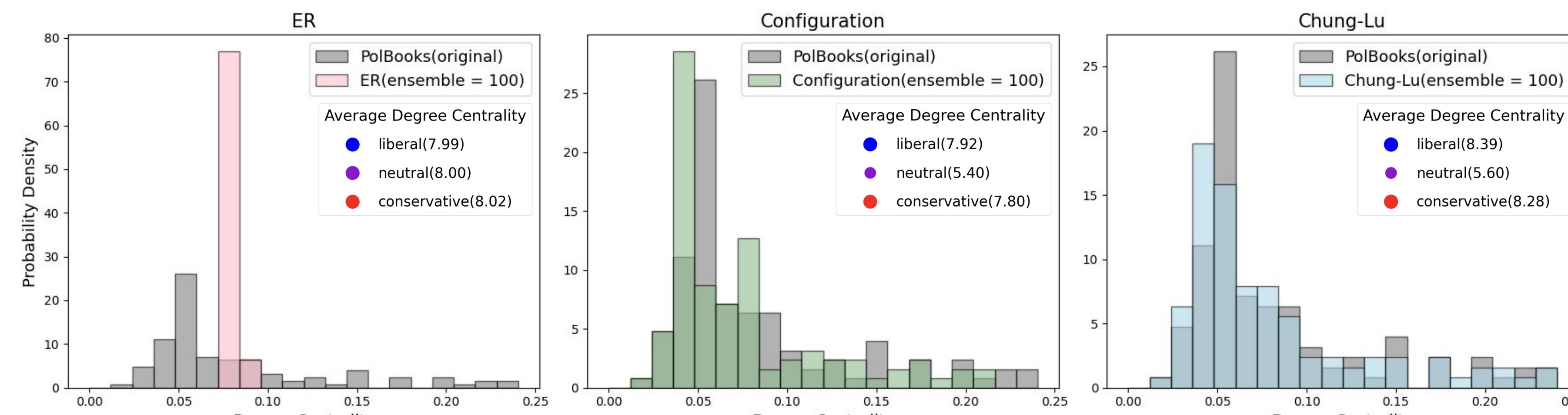
: 네트워크 내에서 각 차수값을 가진 노드의 빈도 또는 확률을 나타내는 분포
무작위 모델에서 유지되는 특성에 따라 원본 차수 분포와의 유사도가 그대로 드러남.

1. ER 모델: 차수 분포 유지 X
두꺼운 꼬리가 없는 포아송 분포 형태를 보임.
2. Configuration 모델: 차수 분포 그대로 유지
원본 차수 분포와 거의 동일한 형태를 보임.
3. Chung-Lu 모델: 차수 분포 확률적 유지
이상치를 제외하고는 두꺼운 꼬리 분포 등 비슷한 형태를 보임.

Centrality ★

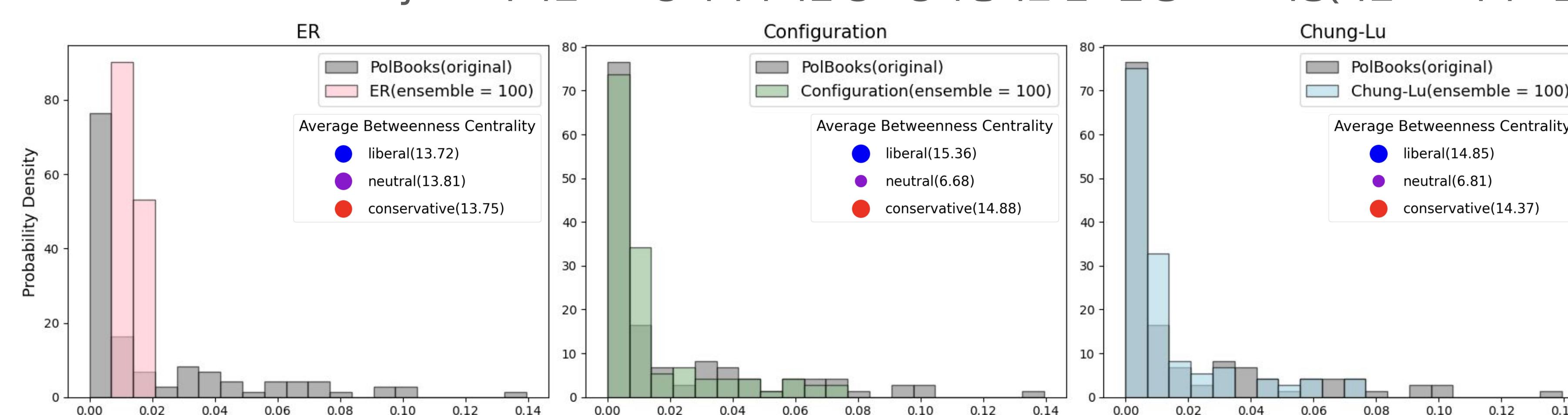
중심성(Centrality): 네트워크 안에서 특정 노드가 얼마나 중요한지를 나타내는 척도

- Degree Centrality: 노드의 차수(이웃수)를 중요도로 사용



* 본 분석에서는 전체 중심성 분포와 도서 유형별 중심성 평균값을 확인함.

- Betweenness Centrality: 노드가 다른 노드 쌍 사이의 최단 경로 상에 놓이는 빈도를 중요도로 사용(다른 노드까지 도달하는데 더 많이 사용된 노드일수록 중요함)



원본 모사 여부

	중심성 분포 (hist)		도서 유형별 평균	
	dgr	btw	dgr	btw
ER	X	X	X	X
Config	O	O	O	X
Chung-Lu	O	O	O	X

1. 엣지 연결 확률 p 가 상수인 ER 모델은 중심성이 전반적으로 동일.
2. 차수를 유지한 모델 (Config/Chung-Lu)이 중심성 분포도 잘 유지함.
3. 차수와 밀접하게 연관된 Degree Centrality는 유형별 평균값도 유지됨.
4. Betweenness Centrality의 유형별 평균값은 반대로 나타남.
→ 원본 네트워크에서 liberal과 conservative를 이어주는 neutral 유형은 차수가 낮았기 때문에 무작위 연결에서 Betweenness Centrality가 급격히 낮아짐.

Conclusion

- 원본 네트워크에서는 conservative와 liberal의 분리가 나타나며 neutral은 그 사이를 잇는 구조가 뚜렷이 드러남.
- 무작위 네트워크에서는 차수를 유지하지 않는 ER 모델은 대부분의 특성을 유지하지 못함.
반면, 차수를 유지하는 Configuration 모델과 Chung-Lu 모델은 차수 분포와 degree centrality를 잘 모사하였으나, betweenness centrality는 분포만 유사하게 유지하고 neutral 유형이 liberal과 conservative 집단을 이어주는 양상은 유지하지 못한 것을 확인함.
- 앞서 관찰된 세 유형 간의 구조적 특성이 무작위 네트워크에서는 유지되지 않는 원본 네트워크만의 고유한 특성이며, 차수 분포에만 의존하는 특성은 아님을 알 수 있음.

Members

분석 파트별 담당자는 기호로 표시. 아래의 QR을 통해 무작위 그래프 생성 패키지 사용 가능 김시현 ★ 이해정 ♣ 장해린 ↔

