1 Community detection in sparse random networks

1.1 Rahmenbedingungen und Hilfslemmata

Sei $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ ein ungerichteter Graph ($\mathcal{V}$ sei die Menge der Ecken, oBdA. $\mathcal{V} = \{1, ..., N\}$, $\mathcal{E}$ die Menge der Kanten). Sei $\mathbf{W} = (W_{i,j}) \in \{0,1\}^{N \times N}$ die zugehörige Adjazenzmatrix, mit $W_{i,j} = 1 \iff (i,j) \in \mathcal{E}$ ($\mathbf{W}$ symmetrisch) und $W_{i,i} = 0$, $i \in \{1, ..., N\}$.
Unter der Nullhypothese ist der Graph $\mathcal{G}$ eine Realisierung von $\mathbb{G}(N, p_0)$, Erdős–Rényi-Graph mit N Ecken und Verbindungswahrscheinlichkeit $p_0 \in (0,1)$ ($\iff$ Die Einträge über der Diagonalen von $\mathbf{W}$ sind iid verteilt mit $\mathbb{P}(W_{i,j} = 1) = p_0 \forall i < j$). Außerdem sei p_0 bekannt.
Unter der Alternative existiert eine Teilmenge $S \subset \mathcal{V}$, so dass $\mathbb{P}(W_{i,j} = 1) = p_1 \forall i,j \in S, i \neq j$ mit $p_1 > p_0$. Die Teilmenge S ist nicht bekannt, es wird aber angenommen, dass $|S| = n$ bekannt ist. Bezeichne H_0 die Nullhypthese, es wird H_0 gegen $H := \cup_{|S|=n} H_S$ getestet. Es wird asymptotisches Verhalten betrachtet, mit
$N \to \infty, n = n(N) \to \infty, n/N \to 0, n/\log(N) \to \infty$.
Weiterhin hängen auch $p_0 = p_0(N)$ und $p_1 = p_1(N)$ von N ab. Ein Test ϕ erhalte $\mathbf{W}$ als Eingabe, so dass $\phi(\mathbf{W}) = 1$, falls eine obige Teilmenge S existiert und $\phi(\mathbf{W}) = 0$ andernfalls. Das Risiko eines Tests ϕ sei

$$\gamma_N(\phi) = \mathbb{P}_0(\phi = 1) + \max_{|S|=n} \mathbb{P}_S(\phi = 0).$$

Eine Folge (ϕ_N) von Tests für eine Folge $\mathbf{W}_N$ von Adjazenzmatrizen heißt asymptotisch schwach, bzw. stark, falls $\gamma_N(\phi_N(\mathbf{W}_N)) \to 0$, bzw. $\gamma_N(\phi_N(\mathbf{W}_N)) \to 1$. Die Hypothesen verschmelzen asymptotisch, falls

$$\gamma_N^* := \inf_{\phi} \gamma_N(\phi) \to 1$$

und trennen sich vollständig asymptotisch, falls

$$\gamma_N^* \to 0.$$

Wir definieren:
$\lambda_0 := Np_0$, $\lambda_1 := np_1$
$W_n^* := \max_{|S|=n} W_S$, $W_{k,S}^* := \max_{T \subset S, |T|=k} W_T$, $W_S := \sum_{i,j \in S, i<j} W_{i,j}$, $W_n^\dagger := \sup_{k=n/u_N}^{n} \frac{W_k^*}{k}$, $u_N := \log(\log(N/n))$
Wir nehmen an, dass $N^2 p_0 \to \infty$, $n^2 p_1 \to \infty$. Definiere weiter
$\alpha := \frac{\log(\lambda_0)}{\log(N/n)}$, $H_p(q) := q \log(q/p) + (1-q)\log(\frac{1-q}{1-p})$, $p, q \in (0,1)$.

Lemma 1.1 (Chernoffs bound) *Für $n \in \mathbb{N}$, $q, p \in (0,1)$ gilt:*

$$\mathbb{P}(Bin(n,p) \geq qn) \leq \exp(-nH_p(q)).$$

Lemma 1.2 *Sei $h(x) = x\log(x) - x + 1$. Für $0 < p \leq q < 1$ gilt:*

$$0 \leq H_p(q) - ph(q/p) \leq O(\frac{q^2}{1-q}).$$

Lemma 1.3 *Für $1 \leq k \leq n$ mit $k, n \in \mathbb{N}$ gilt*

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

Lemma 1.4 *Für jedes $\lambda > 1$ bezeichne η_λ den kleinsten Wert, der die Gleichung $\eta = \exp(\lambda(\eta - 1))$ löst. Sei C_m die größte Zusammenhangskomponente in $\mathbb{G}(m, \lambda/m)$ und nehme an, dass $\lambda > 1$ fest sei. Dann gilt*

$$|C_m| \sim^{\mathbb{P}} (1-\eta_\lambda)m$$
$$W_{C_m} \sim^{\mathbb{P}} \frac{\lambda}{2}(1-\eta_\lambda^2)m.$$

1.2 Resultate

Satz 1.5 (Broad Scan Test) *Der Scan Test basierend auf $W_n^\dagger$ ist asymptotisch stark, falls*
(i) $\limsup \alpha \leq 1$ *und* $\liminf (1-\alpha) \sup_{k=n/u_N}^{n} \frac{\mathbb{E}_S[W_{k,S}^*]}{k} > 1$;
oder
(ii) $\alpha \to 0$ *und* $\liminf \lambda_1 > 1$.

Satz 1.6 *Angenommen,* $\zeta := \frac{(\lambda_1 - \lambda_0 n/N)^2}{\lambda_0} \frac{n^2}{N} \to 0$*. Dann sind alle Tests asymptotisch schwach in jedem der fogenden Szenarien:*
(i) $\lambda_0 \to 0$, $\lambda_1 \to 0$, $\limsup \frac{I_{\lambda_0}}{I_{\lambda_1}} \frac{\log(n)}{\log(N)} < 1$, *wobei* $I_\lambda := \lambda - 1 - \log(\lambda)$*;*
(ii) $0 < \liminf \lambda_0 \leq \limsup \lambda_0 < \infty, \lambda_1 \to 0$*;*
(iii) $\lambda_0 \to \infty$ *mit* $\alpha \to 0, \limsup \lambda_1 < 1$*;*
(iv) $0 < \liminf \alpha \leq \limsup \alpha < 1, \limsup(1-\alpha) \sup_{k=n/u_N}^{n} \frac{\mathbb{E}_S[W_{k,S]^*]}}{k} < 1$.

Der Beweis des letzten Resultates ist fehlerhaft.