7.1 Planted partition model and overview

In the second part of the course, we will study the problem of *community detection* in a broad sense. Consider the following abstract *planted partition model*, where a matrix $A = (A_{ij})_{1 \leq i < j \leq n}$ is observed whose distribution depends on the latent labels $\sigma = (\sigma_1, \ldots, \sigma_n) \in \{\pm\}^n$, such that

$$A_{ij} \sim \begin{cases} P & \sigma_i = \sigma_j \\ Q & \sigma_i \neq \sigma_j \end{cases}$$

Given A, the goal is to recover the labels σ accurately.

Two prominent special cases are the following:

Stochastic block model (SBM) Here $P = \text{Bern}(p)$ and $Q = \text{Bern}(q)$. In this case the set of vertices $[n]$ is partitioned into two communities $V_+ = \{i : \sigma_i = +\}$ and $V_- = \{i : \sigma_i = +\}$, and A is the adjacency matrix of a random graph, such that two nodes i and j are connected with probability p if they belong to the same community, and with probability q if otherwise. The case of $p > q$ is referred to as "assortative" and $p < q$ as "disassortative".

The community structure is determined by the vector σ, which, depending on the problem formulation, could either be fixed or random. We will frequently consider special cases:

- iid model: Each σ_i is equally likely to be $\pm$ (Rademacher) and independently.
- exact bisection: $|V_+| = |V_-| = n/2$ (when n is even) and the partition is chosen uniformly at random from all bisections.

Typically these two models behave very similarly.

Spiked Wigner model (Rank-one deformation) Here $P = N(\sqrt{\frac{\lambda}{n}}, 1)$ and $Q = N(-\sqrt{\frac{\lambda}{n}}, 1)$. In matrix notation,

$$A = \sqrt{\frac{\lambda}{n}} \sigma \sigma^\top + Z \tag{7.1}$$

where Z is such that $\{Z_{ij} : 1 \leq i < j \leq n\}$ are iid $N(0, 1)$. Therefore A can be viewed is a rank-one perturbation of a Gaussian Wigner matrix.

As opposed to the treatment of the planted clique problem in Part **??**, we will be focusing on

- Sharp threshold, i.e., finding the exact constant in the fundamental limit (and achieving them with fast algorithms).

- "Sparse" graphs, where the edge density tends to zero (at different speed), unlike the hidden clique model $G(n, \frac{1}{2}, k)$

We will focus on the following three formulations (recovery guarantees):

Detection Here there is a null model. For example,

- For spiked Wigner model, the null hypothesis is A is iid Gaussian. The sharp threshold is given by $\lambda = 1$, in the sense that for any fixed ϵ, it is possible to test the hypotheses with vanishing error probability if $\lambda \geq 1 + \epsilon$, and impossible if $\lambda \leq 1 - \epsilon$.
- For SBM with bisection, we want to test against the null hypothesis of no community structure, that is, an Erdős-Rényi graph $G(n, \frac{p+q}{2})$ with the same average degree. The most interesting regime is bounded average degree $p = \frac{a}{n}, q = \frac{b}{n}$ for constants a, b, and the sharp threshold is given by $\frac{(a-b)^2}{2(a+b)} = 1$.

Correlated (weak) recovery Here and below, there is no null model. The goal is to recovery the community structure (labels) better than random guessing. Let $\hat{\sigma} = \hat{\sigma}(A)$ be the estimator. Its overlap with the true labels σ is $|\langle \hat{\sigma}, \sigma \rangle|$ and the number of misclassification errors (up to a global sign flip) is expressed as

$$\ell(\sigma, \hat{\sigma}) = \min_{\pm} \|\hat{\sigma} \pm \sigma\|_1 = n - |\langle \hat{\sigma}, \sigma \rangle|.$$

In the iid setting, random guessing would yield, by CLT, $|\langle \hat{\sigma}, \sigma \rangle = O_P(\sqrt{n})|$ and $\mathbb{E}[|\langle \hat{\sigma}, \sigma \rangle|] = o(n)$. The goal of weak recovery is to achieve a positive correlation, namely

$$\mathbb{E}[|\langle \hat{\sigma}, \sigma \rangle|] = \Omega(n)$$

Although in general detection and correlated recovery are two different problems, for both SBM and spiked Wigner the thresholds coincide. In fact, for certain models one can have a generic reduction between the problems (e.g. spiked Wigner, see Homework).

(Almost) exact recovery *Almost exact recovery* means achieving a vanishing misclassification rate: $\mathbb{E}\ell(\sigma, \hat{\sigma}) = o(n)$. Typically the sharp threshold is expressed in terms of Hellinger distance as $H^2(P, Q) \gg \frac{1}{n}$.

Exact recovery means $\ell(\sigma, \hat{\sigma}) = 0$ with probability tending to 1. Typically the sharp threshold is given by $H^2(P, Q) = \frac{(2+\epsilon)\log n}{n}$.

A more statistical flavored question is to characterize the optimal (in the sense of minimax) misclassification rate $\frac{1}{n}\ell(\sigma, \hat{\sigma})$, which typically behaves as $\exp(-\frac{H^2(P,Q)}{2})$.

7.2 Detection threshold for SBM

We want to test the hypothesis

$$H_0 : G \sim G(n, \frac{p+q}{2}) \text{ vs. } H_1 : G \sim SBM(n, p, q).$$

Under the SBM model, we assume the the labels $\sigma = (\sigma_1, \ldots, \sigma_n)$ are either iid $\text{Rad}(\frac{1}{2})$, or drawn uniformly at random from all bisections. The detection problem is non-trivial in the regime of bounded average degree:

$$p = \frac{a}{n}, \ q = \frac{b}{n}, \tag{7.2}$$

where a, b are constants.

Theorem 7.1. *If $\frac{(a-b)^2}{2(a+b)} > 1$, detection is possible, in the sense of total variation that*

$$\text{TV}(\text{Law}(G|H_0), \text{Law}(G|H_1)) \to 1 \tag{7.3}$$

If $\frac{(a-b)^2}{2(a+b)} \leq 1$, detection is impossible, in the sense that

$$\text{TV}(\text{Law}(G|H_0), \text{Law}(G|H_1)) \leq 1 - \Omega(1). \tag{7.4}$$

We start with the impossibility part. For non-detection it is enough to show

$$\chi^2(\text{Law}(G|H_0)||\text{Law}(G|H_1)) = O(1). \tag{7.5}$$

Remark 7.1 (Contiguity)**.** Recall the notion of contiguity (of two sequences of probability measures (P_n) and (Q_n)). We say (P_n) is continguous to (Q_n) if for any sequence of events E_n, $Q_n(E_n) \to 0 \implies P(E_n) \to 0$. Contiguity implies non-detection, because for any sequence of tests

$$Q_n(\text{failure}) \to 0 \implies P_n(\text{success}) \to 0$$

which is bad news.

A sufficient condition of continguity is bounded second moment of likelihood, i.e., $\chi^2(P_n \| Q_n) = O(1)$. Indeed, by Cauchy-Schwarz,

$$P_n(E_n) = \mathbb{E}_{Q_n}\left[\frac{P_n}{Q_n}\mathbb{1}\{E_n\}\right] \leq \underbrace{\sqrt{\mathbb{E}_{Q_n}\left[\left(\frac{P_n}{Q_n}\right)^2\right]}}_{\sqrt{\chi^2+1}} Q_n(E_n) \to 0$$

The following lemma is very useful for computing χ^2(mixture distribution||simple distribution). The introduction of two iid copies of randomness is typical in second moment calculation (cf. Section **??**).

Lemma 7.1 (Second moment trick)**.** *Suppose we have a parametric family of distributions $\{P_\theta : \theta \in \Theta\}$. Given a prior on the parameter space Θ, define the mixture distribution:*

$$P_\pi \triangleq \int P_\theta \pi(d\theta).$$

Then we have $\chi^2(P_\pi||Q) = \mathbb{E}G(\theta, \tilde{\theta}) - 1$, where $\theta, \tilde{\theta} \overset{iid}{\sim} \pi$ and $G(\theta, \tilde{\theta})$ is defined by

$$G(\theta, \tilde{\theta}) \triangleq \int \frac{P_\theta P_{\tilde{\theta}}}{Q}$$

.

Proof. The proof is just by Fubini:

$$\begin{aligned}\int \frac{P_\pi^2}{Q} &= \int \frac{(\int P_\theta(x)\pi(d\theta))(\int P_{\tilde\theta}(x)\pi(d\tilde\theta))}{Q(x)}\mu(dx)\\ &= \int \pi(d\theta)\pi(d\tilde\theta)\underbrace{\left(\frac{P_\theta(x)P_{\tilde\theta}(x)}{Q(x)}\mu(dx)\right)}_{G(\theta,\tilde\theta)}\end{aligned}$$

□

Example 7.1 (Gaussian)**.** Consider $P_\theta = N(\theta, I_d)$ and $Q = N(0, I_d)$, and let π be some distribution on $\mathbb{R}^d$. Then $\chi^2(P_\pi\|Q) = \mathbb{E}[\langle\theta,\tilde\theta\rangle] - 1$, where $\theta, \tilde\theta \overset{\text{i.i.d.}}{\sim} \pi$.

The calculation for SBM can be carried out in a very general setting. Consider P and Q in place of Bern(p) and Bern(q). For each label $\sigma \in \{\pm\}^n$, the distribution of the adjacency matrix is

$$P_\sigma = \text{Law}(A|\sigma) = \prod_{i<j}(P\mathbb{1}_{\{\sigma_i=\sigma_j\}} + Q\mathbb{1}_{\{\sigma_i\neq\sigma_j\}}) = \prod_{i<j}\left(\frac{P+Q}{2} + \frac{P-Q}{2}\sigma_i\sigma_j\right) \tag{7.6}$$

and the null distribution is $P_0 = \prod_{i<j}\frac{(P+Q)}{2}$. Fix two assignment $\sigma, \hat\sigma \in \{\pm1\}^n$. Then

$$\begin{aligned}G(\sigma,\hat\sigma) &= \int \frac{P_\sigma P_{\hat\sigma}}{P_0}\\ &= \int\prod_{i<j}\frac{\left(\frac{P+Q}{2}+\frac{P-Q}{2}\sigma_i\sigma_j\right)\left(\frac{P+Q}{2}+\frac{P-Q}{2}\tilde\sigma_i\tilde\sigma_j\right)}{\frac{P+Q}{2}}\\ &= \prod_{i<j}\left[\underbrace{\int\frac{P+Q}{2}}_{=1} + \underbrace{\int\frac{P-Q}{2}}_{=0}\sigma_i\sigma_j + \underbrace{\int\frac{P-Q}{2}}_{=0}\hat\sigma_i\hat\sigma_j + \underbrace{\int\frac{(P-Q)^2}{2(P+Q)}}_{\triangleq\frac{\tau}{2n}}\sigma_i\sigma_j\hat\sigma_i\hat\sigma_j\right]\\ &= \prod_{i<j}\left[1+\rho\sigma_i\sigma_j\hat\sigma_i\hat\sigma_j\right]\\ &\leq \exp\left(\rho\sum_{i<j}\sigma_i\sigma_j\hat\sigma_i\hat\sigma_j\right) \leq \exp\left(\frac{\rho}{2}\langle\sigma,\hat\sigma\rangle^2\right)\end{aligned}$$

Thus, by Lemma **??**, we have

$$\chi^2(P_1\|P_0) + 1 = \mathbb{E}_{\sigma,\tilde\sigma}\left[\exp\left(\frac{\rho}{2}\langle\sigma,\hat\sigma\rangle^2\right)\right]$$

where $\tilde\sigma$ is an iid copy of σ.

For SBM(n,p,q), under the scaling (**??**), we have

$$\rho = \frac{\tau+o(1)}{n}, \quad \tau \triangleq \frac{(a-b)^2}{2(a+b)}.$$

Next we consider two situations:

Random labels: $\sigma, \hat\sigma \overset{iid}{\sim} \{\pm 1\}^n$. By CLT, $\frac{1}{\sqrt{n}}\langle \sigma, \hat\sigma\rangle = \frac{1}{\sqrt{n}}\sum_{i=1}^n \sigma_i \hat\sigma_i \xrightarrow{\mathrm{D}} Z \sim N(0,1)$. Assuming convergence of MGF (see Lemma **??** next), we have

$$\begin{aligned}
\chi^2(P_1 \| P_0) + 1 &= \mathbb{E}\exp\left(\frac{\tau + o(1)}{2n}\langle \sigma, \hat\sigma\rangle^2\right) \\
&\to \mathbb{E}\left(\frac{\tau + o(1)}{2} Z^2\right) \\
&= \begin{cases} \infty & \text{if } \tau \geq 1 \\ \text{constant} & \text{if } \tau < 1. \end{cases}
\end{aligned} \tag{7.7}$$

Exact bisection: Let us consider the case where $\sigma, \hat\sigma$ are drawn iid and uniformly at random from the set $\{\theta \in \{\pm 1\}^n : \sum \theta_i = 0\}$. For simplicity, write

$$\sigma = 2\xi - \mathbf{1}, \quad \tilde\sigma = 2\tilde\xi - \mathbf{1},$$

Then $\langle \sigma, \hat\sigma\rangle = 4\langle \xi, \hat\xi\rangle - n$. Both $\xi, \tilde\xi$ are iid uniform random $\frac{n}{2}$-sparse binary vectors. So

$$\langle \xi, \hat\xi\rangle \sim \text{Hypergeometric}(n, \frac{n}{2}, \frac{n}{2}),$$

which means (check!)[1]

$$\frac{\langle \xi, \hat\xi\rangle - \frac{n}{4}}{\sqrt{\frac{n}{16}}} \xrightarrow{\mathrm{D}} Z \sim N(0,1).$$

Thus the dichotomy (**??**) applies to bisection as well.

To pass from weak convergence to convergence of the MGF, the following lemma is useful:

Lemma 7.2 (Convergence of MGF)**.** *Assume that* $X_n \xrightarrow{\mathrm{D}} X$. *Let* $M_n(t) = \mathbb{E}\exp(tX_n)$ *and* $M(t) = \mathbb{E}\exp(tX)$. *If there exists some constant* $\alpha > 0$ *such that*

$$\sup_n P(|X_n| > x) \leq \exp(-\alpha x)$$

for all $x > 0$, *then* $M_n(t) \to M(t)$ *for all* $|t| < \alpha$.

Remark 7.2.
- The critical case of $\frac{(a-b)^2}{2(a+b)} = 1$ also implies non-detection. Proving this is outside the scope of this section as the χ^2 truly blows up.
- The threshold of the spiked Wigner model (**??**) is given by $\lambda = 1$. This can be proved by the same second moment method (homework).

7.3 Test by counting cycles

Below we describe a test for

$$H_0 : G \sim G(n, \frac{p+q}{2}) \text{ vs. } H_1 : G \sim SBM(n, p, q).$$

[1]Note that the variance of Hypergeometric$(n, \frac{n}{2}, \frac{n}{2})$ is exactly half of its counterpart Binom$(\frac{n}{2}, \frac{1}{2})$. Why? Think about sampling with and without replacements.

that achieves the sharp threshold in Theorem **??**, following **?**. We will consider the labels being iid $\text{Rad}(\frac{1}{2})$. The test is based on counting "short" cycles – by short we mean much shorter than the longest cycle, but the length still need to be slowing growing. As there is no generic polynomial-time algorithm for counting k-cycles (C_k) for growing k, in the next section we make it polynomial-time relying on the randomness of the graph.

Consider the number of k-cycles (not induced cycles) as the test statistic, denoted by X_k. As a warmup, consider the behavior of X_k in $G(n, \frac{d}{n})$. Then by union bound,

$$\mathbb{P}(X_k > 0) \leq \mathbb{E}[X_k] = \binom{n}{k} k! \frac{1}{2k} \left(\frac{d}{n}\right)^k \leq d^k,$$

where the overcounting factor $2k$ is the number of symmetries (automorphisms) of C_k, namely, cyclic shift and flip. Thus there are no cycles of growing length if $d < 1$. Of course, this first-moment calculation does not tell us about existence. Nevertheless it is known that if $d \geq 1$, the longest cycle is of length $\Omega(n)$ (**?**, Chap. 8).

Now let's get back to the original problem of testing $G(n, \frac{a+b}{2n})$ versus $SBM(n, \frac{a}{n}, \frac{b}{n})$. Assume that $a > b$. Define

$$s = \frac{a-b}{2}, \quad d = \frac{a+b}{2}.$$

The threshold is then given by $s^2 \geq d$. Since $d > s$, this implies $s > 1$ and $a > 2$.

<u>Intuition</u>: For k not too big, X_k has a Poisson limit under both model with different parameters. To prove the success of the test (based on thresholding X_k), it suffices to compute its mean and variance. We will show

$$\begin{aligned} &\text{Under } H_0\text{:} \quad \mathbb{E}X_k \approx d^k, \quad \text{Var}X_k \leq d^k, \\ &\text{Under } H_1\text{:} \quad \mathbb{E}X_k \approx d^k + s^k, \quad \text{Var}X_k \leq d^k \end{aligned}$$

Under the condition $s^2 > d$, we have

$$\mathbb{E}_1[X_k] - \mathbb{E}_0[X_k] \gg \sqrt{\text{Var}_0(X_k) + \text{Var}_1(X_k)}$$

as k growing, and hence the test $\mathbb{1}\left\{X_k \leq d^k + \frac{s^k}{2}\right\}$ succeeds.

7.3.1 First moment calculation

Under H_0. First we note that

$$X_k = \frac{1}{2k} \sum_{\substack{v_1,\ldots,v_k; \\ \text{all ordered } k\text{-tuple} \\ \text{from } V(G)}} \mathbb{1}_{\{v_1 \sim v_2, v_2 \sim v_3, \ldots, v_k \sim v_1\}},$$

which implies

$$\mathbb{E}X_k = \frac{1}{2k} \underbrace{\binom{n}{k} k!}_{\triangleq [n]_k} \underbrace{\mathbb{P}\{v_1 \sim v_2, v_2 \sim v_3, \ldots, v_k \sim v_1\}}_{=(\frac{d}{n})^k} \approx \frac{1+o(1)}{2k} d^k \tag{7.8}$$

under H_0, where the last equality holds provided $k = o(\sqrt{n})$ (Why? Think about birthday problem).

Under H_1. We just need to recompute the probability in (**??**), which now depends on the labels of the vertices. Consider the adjacency matrix A. Then given any two vertices v_i, v_{i+1}, we have

$$A_{v_i,v_{i+1}} \sim \begin{cases} \text{Bern}(p) \text{ if } \sigma_i = \sigma_{i+1} \\ \text{Bern}(q) \text{ if } \sigma_i \neq \sigma_{i+1}. \end{cases}$$

Given any k-tuple $\{v_1, v_2, \ldots, v_k\}$ of vertices, suppose N denotes the number of disagreements of adjacent labels, given by

$$N = \sum_{i=1}^{k} \mathbb{1}_{\{\sigma(v_i) \neq \sigma(v_{i+1})\}}$$

with $k+1$ understood as 1 circularly. Write

$$N = \underbrace{\sum_{i=1}^{k-1} \mathbb{1}_{\{\sigma(v_i) \neq \sigma(v_{i+1})\}}}_{\triangleq T} + \underbrace{\mathbb{1}_{\{\sigma(v_k) \neq \sigma(v_1)\}}}_{\triangleq S}$$

Then we have $T \sim \text{Binom}(k-1, \frac{1}{2})$ and

$$S = \begin{cases} 0 & T \text{ is even} \\ 1 & T \text{ is odd} \end{cases}$$

is a parity bit, so that $N = S + T$ is always even.

It is clear that conditioned on $N = m$, the probability of $v_1, \ldots, v_k$ forming a cycle is $q^m p^{k-m}$. Note that

$$\mathbb{P}(N = m) = \begin{cases} 0 & m \text{ odd} \\ \mathbb{P}(\text{Binom}(k-1, \frac{1}{2}) = m-1 \text{ or } m) = \binom{k}{m} 2^{-k+1} & m \text{ even} \end{cases}$$

Thus

$$\begin{aligned} \mathbb{P}(v_1 \sim v_2 \sim \cdots \sim v_k) &= \sum_{m=0}^{k} q^m p^{k-m} \cdot \mathbb{P}(N = m) \\ &= \sum_{\substack{m=0 \\ m \text{ even}}}^{k} q^m p^{k-m} \binom{k}{m} 2^{-k+1} \\ &= \sum_{m=0}^{k} \frac{(-q)^m p^{k-m} + q^m p^{k-m}}{2} \binom{k}{m} 2^{-k+1} \\ &= \left(\frac{p+q}{2}\right)^k + \left(\frac{p-q}{2}\right)^k = n^{-k}(s^k + d^k). \end{aligned}$$

Thus, under H_1,

$$\begin{aligned} \mathbb{E}(X_k) &= \frac{[n]_k}{2k} \left\{ \left(\frac{p+q}{2}\right)^k + \left(\frac{p-q}{2}\right)^k \right\} \\ &\overset{k=o(\sqrt{n})}{=} \frac{1+o(1)}{2k} \left(s^k + d^k\right). \end{aligned}$$

7.3.2 Variance analysis

We only consider the variance under the null, as the alternative is similar. Given ordered k-tuple of vertices $T = (v_1, \ldots, v_k)$, define $b_T = \mathbb{1}\{v_1 \sim v_2, \cdots, v_k \sim v_1\}$. Then under H_0, we have

$$\mathrm{Var}(X_k) = \frac{1}{4k^2} \sum_{T,T'} \mathrm{Cov}(b_T, b_{T'}) = \frac{1}{4k^2} \left(\sum_T \underbrace{\mathrm{Var}(b_T)}_{\leq \mathbb{E}[b_T] = d^k} + \sum_{\substack{T \neq T' \\ T \cap T' \neq \emptyset}} \mathrm{Cov}(b_T, b'_T) \right).$$

Consider two distinct k-cycles T and T' that are overlapping. Let

$$\ell = \text{number of common edges}, \quad v = \text{number of common vertices}.$$

Note that

- $\mathrm{Cov}(b_T, b_{T'}) \leq \mathbb{E}[b_T b_{T'}] = p^{2k-\ell}$
- Crucially,

  $$v \geq \ell + 1.$$

  This is because the intersection of two cycles is a forest (each connected component is a path), so that $v = \ell + \mathrm{cc}$.

Combining all this, we get

$$\begin{aligned}
\sum_{\substack{T \neq T' \\ T \cap T' \neq \emptyset}} \mathrm{Cov}(b_T, b'_T) &\leq \sum_{\ell=1}^{k-1} [n]_{2k-v} [k]_v \left(\frac{d}{n}\right)^{2k-\ell} \\
&\overset{v \geq \ell+1}{\leq} \sum_{\ell=1}^{k-1} n^{2k-\ell-1} k! \left(\frac{d}{n}\right)^{2k-\ell} \\
&\leq \frac{1}{n} k^{k+1} d^{2k} \\
&= o(1), \quad \text{provided that } k = o(\log n / \log\log n).
\end{aligned}$$

So we get

$$\mathrm{Var}(X_k) = \frac{1}{4k^2} d^k + o(1).$$

7.4 Approximately counting cycles in polynomial time

<u>A caveat:</u> The naive way of counting (exhaustive search) k-cycles takes n^k time, which is not polynomial in n if $k \to \infty$. From the previous analysis, we see that we need to count k-cycles with slowly growing k.

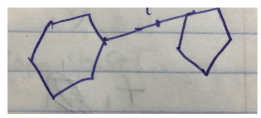
<u>Fix:</u> The trick is to use the sparsity of the random graph and approximately count the number of k-cycles.

Definition 7.1 (ℓ-tangle free)**.** An ℓ-tangle is a connected subgraph of diameter at most 2ℓ that contains atleast two cycles.

A graph G is called ℓ-tangle free if no subgraph of G is an ℓ-tangle. In other words, for all $v \in V(G)$, its ℓ-hop neighborhood $N_\ell(v)$ contains at most one cycle.

Lemma 7.3. *If $G \sim G(n, \frac{d}{n})$ and d is a constant, then G is ℓ-tangle free if $\ell = o(\log n)$ (In general $\ell \log d = c \log n$ for small constant c suffices).*

Proof. Suppose G contains an ℓ-tangle. Then G must contain a subgraph of the following form



with m edges and v vertices, such that $m \leq 4\ell$ and $m \geq v + 1$. Then by union bound, such a graph exists with probability

$$\leq n^v \left(\frac{d}{n}\right)^m \leq \frac{d^{O(\ell)}}{n} \to 0,$$

when $\ell \log d \ll \log n$. □

Next we discuss the connection between counting and linear algebra. Let's start with triangles ($k = 3$):

Example 7.2 (Counting triangles)**.** Suppose that A is the adjecency matrix of G. Given any vertex v in G,

$$(A^3)_{vv} = \sum_{a,b} A_{va} A_{ab} A_{bv}$$

is in fact twice the number of triangles incident to v. Therefore, $\mathsf{Tr}(A^3) = 6\times$ the number of triangles in G.

To count find k-cycles one can consider computing $\mathsf{Tr}(A^k)$, which can be done in the time of eigenvalue decomposition. But

$$\mathsf{Tr}(A^k) = \text{ number of closed walks of length } k \gg \text{ number of } k\text{-cycles} .$$

The strategy next is use the tangle-free structure and count the number of *non-backtracking* (NB) paths.

Definition 7.2 (Non-backtracking walk)**.** We say

- $(v_1, v_2, \ldots, v_k)$ is a NB walk if $v_t \sim v_{t+1}$ and $v_t \neq v_{t-2}$ for all t.
- $(v_1, v_2, \ldots, v_k)$ is a NB cycle if $v_t \sim v_{t+1}$ and $v_t \neq v_{t-2}$ for all t and $v_1 = v_k$.

For example,

Consequences: Conditioned on G being $2k$-tangle free, any NB cycle of k steps is either a k-cycle, or an m cycle traversed for $\frac{k}{m}$ times. Otherwise, we have a $2k$-tangle such as two short cycles

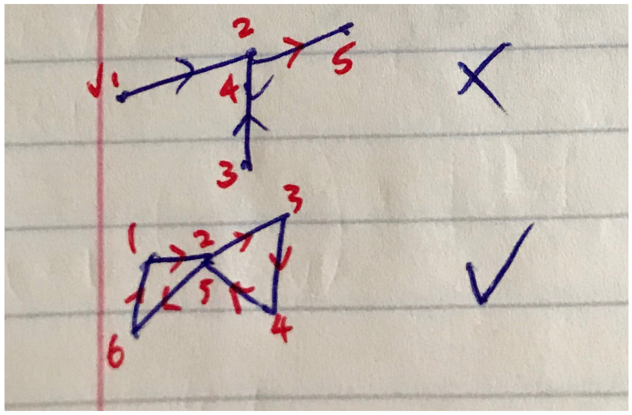


Figure 7.1: Examples of backtracking and non-backtracking.

sharing a vertex (see Fig. ?? above). This reduces the problem to counting the number of NB cycles of length m, for all $m = 1, \dots, 2k$.

It is easy to count NB walk of length k recursively: Let $N_{uv}^m = \#$ of NB walks from $u \to v$ for k steps. Then our goal is expressed as

$$\sum_{v \in V(G)} N_{vv}^m$$

So it suffices to compute N_{uv}^m for all pairs u, v.

It turns out N_{uv}^m is given by the following three-term recursion:

$$N_{uv}^{m+1} = \sum_{w \sim v} N_{uw}^m - (d_v - 1) N_{uv}^{m-1}. \tag{7.9}$$

In matrix notation: let $N^{(m)} = (N_{uv}^m)$ and $D = \mathrm{diag}(d_v)$. Then we have[2]

$$\begin{cases} N^{(m+1)} = N^{(m)} \cdot A - N^{(m-1)}(D - \mathbf{I}), \\ N^{(1)} = A, \quad N^{(2)} = A^2 - D \end{cases} \tag{7.10}$$

which means we can compute all N_{uv}^m's using matrix multiplication.

Finally, to justify (??), simply notice that the first term on the RHS counts all NB walks of m steps from u to a neighbor w of v, which, followed by another step from w to v, constitute a walk of $m+1$ steps from u to v. But, it can be backtracking. So we need to subtract those out, and that it precisely the second term: fix any NB walk from u to v of $m-1$ steps, say, $u, \dots, v', v$, where $v' \in N(v)$. Append this walk by $w \in N(v) \backslash \{v'\}$ constitutes a NB walk from u to w in m steps.

[2] In the special case of d-regular graphs, (??) becomes $N^{(m+1)} = N^{(m)} \cdot A - (d-1)N^{(m-1)}$. This means $N^{(m)}$ is a polynomial of A, in fact, the Chebyshev polynomial, which satisfies the same three-term recurrence. See ? for more.