Lecture 3: Overview of path homology of digraphs

Mohabat

March 11, 2022

Contents

1 Homological dimension

- in this section, $\mathbb{K} := \mathbb{F}_2$

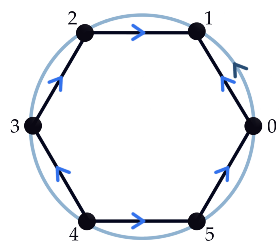
Definition 1.1. The **homological dimension** of a digraph G is

$$\dim_h G := \sup\{k : |H_k(G)| > 0\}$$

1.1 Some examples

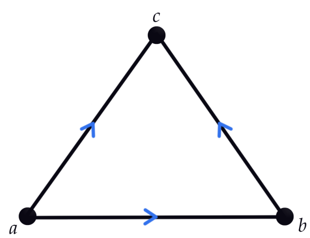
1. G: a polygon (i.e., a cyclic digraph)
   - if G is neither a triangle nor square, then $|H_1| = 1$ and $|H_p| = 0$ for $p \geq 2$ so that

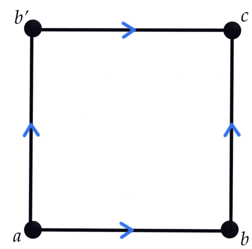
$$\dim_h G = 1$$



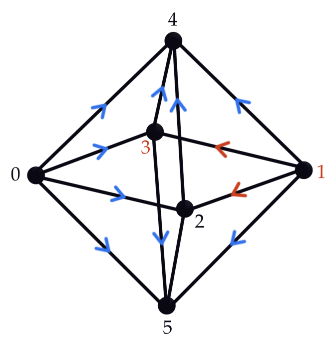
- if G is either a triangle or a square, then $|H_p| = 0$ for $p \geq 1$ and hence

$$\dim_h G = 0$$





2. Let G be the octahedron here:



   - $|H_2| = 1$
   - $|H_p| = 0$ for $p \geq 3$
   - thus, $\dim_h G = 2$

3. there *are* finite digraphs with $\dim_h G = \infty$ as the one here:
   - constructed by Gabor Lippner and Paul Horn in 2012

1.2 Random digraphs

- we are interested in the *homological dimension* of a randomly generated digraph G
- fix a finite set of vertices $\{1, 2, \ldots, V\}$
- fix $p, q > 0$ with $p + q \leq 1$

- set of arrows in G is defined as:
  - for any 2 vertices $a < b$ there is:
    - an arrow $a \to b$ with probability p;
    - an arrow $b \to a$ with probability q; and
    - no arrow with probability $1 - p - q$
- the so-constructed probability measure on digraphs will be denoted $\mathbb{P} = \mathbb{P}_{p,q,V}$
- randomly generated digraph with $p = q = 0.37, V = 15, E = 86$
  - for this digraph, $\dim_h G = 6$
- set $r := p + q$
- the number E of arrows is random, and it is easy to compute:

$$\mathbb{E}(E) = \frac{r}{2}V(V-1)$$
$$\operatorname{Var}(E) = \frac{1}{2}r(1-r)V(V-1)$$

Definition 1.2. The **degree of a digraph** is the average outgoing degree of the vertices:

$$D = \deg G := \frac{E}{V}$$

- for example: for the above digraph, $D = \frac{86}{15} \approx 5.7$
- for random digraphs, it follows from the above expected value and variance computations, that

$$\mathbb{E}(D) = \frac{r}{2}(V-1)$$
$$\operatorname{Var}(D) = \frac{1}{2}r(1-r)\frac{V-1}{V}$$

- moreover, applying the CLT to the sum of indicators of arrows, we obtain:

$$D_{\text{norm}} := \frac{D - \frac{r}{2}(V-1)}{\sqrt{\frac{1}{2}r(1-r)\frac{V-1}{V}}} \xrightarrow{\mathcal{D}} \text{Normal}(0,1) \text{ as } V \to \infty$$

Proposition 1.3. *If $p + q > 0$, then*

$$\lim_{V\to\infty} \mathbb{P}_{p,q,V}(G \textit{ is connected}) = 1$$

That is,

$$\mathbb{P}_{p,q,V}(\beta_0(G) = 1) \to 1 \textit{ as } V \to \infty$$

1.3 Homological dimension and degree

- turns out that $\dim_h G$ for random digraphs is closely related to the *degree* $D = \frac{E}{V}$
- in > 1,000 samples of randomly generated digraphs, we have observed the following dichotomy:

  Observation: With high probability, either $\dim_h G = 0$ or $\dim_h \asymp D$ (i.e., the dimension is 0 or **asymptotic** to the degree D).

- consider random variables

$$Q := \frac{\dim_h G}{D}$$

$$Q_+ := (Q \mid Q > 0) \text{ (a conditioned RV)}$$

- assume that $p = q \in \left(0, \frac{1}{2}\right)$

Conjecture: There exist positive limits

$$\mu(p) = \lim_{V \to \infty} \mathbb{E}_{p,p,V}(Q_+) \qquad \text{and}$$
$$\tau^2(p) = \lim_{V \to \infty} \mathrm{Var}_{p,p,V}(Q_+) = \lim_{V \to \infty} \mathbb{E}_{p,p,V}(Q_+^2) - \mu(p)^2$$

Besides, we have

$$\mu(p) > 3\tau(p)$$

- plot of empirical functions $\mu(p)$ and $\tau(p)$ computed using the averages of Q_+ and Q_+^2 among all available samples

Conjecture 3.3: We have $Q_+ \xrightarrow{\mathcal{D}} \frac{1}{Z}\mathrm{Normal}_+(\mu, \tau^2)$ as $V \to \infty$ where $\mu := \mu(p)$ and $\tau := \tau(p)$.

That is, for any $x \geq 0$,

$$\lim_{V \to \infty} \mathbb{P}_{p,p,V}(Q_+ \leq x) = \frac{1}{Z} \int_0^x \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{(y-\mu)^2}{2\tau^2}\right) dy$$

where Z is a normalizing factor.

- as one sees in the plot, $\mathbb{P}(0.4 \leq Q_+ \leq 1) \approx 0.9$, i.e.,

$$\mathbb{P}(0.4D \leq \dim_h(G) \leq D \mid \dim_h(G) > 0) \approx 0.9$$

1.4 Homologically trivial and spherical digraphs

Definition 1.4. A digraph G is **homologically trivial** if $\dim_h G = 0$.

- i.e., $\beta_k(G) = 0$ for all $k \geq 1$

Conjecture 3.4: The following limit exists and is positive:

$$T(p) = \lim_{V\to\infty} \mathbb{P}_{p,p,V} \qquad (G \text{ is homologically trivial})$$

Consequently,

$$\frac{\dim_h G}{D} = Q \xrightarrow{\mathcal{D}} T(p)\delta_0 + \frac{1}{Z(1-T(p))}\text{Normal}_+(\mu, \tau^2) \text{ as } V \to \infty$$

Definition 1.5. A digraph G is **homologically spherical of dimension** n if

$$\beta_k(G) = \begin{cases} 1 & \text{if } k = 0, n \\ 0 & \text{else} \end{cases}$$

- in this case, $\dim_h G = n$
- notice that any homologically trivial digraph is also spherical of dimension 0.

Conjecture 3.5: The following limit exists and is positive:

$$S(p) = \lim_{V\to\infty} \mathbb{P}_{p,p,V} \qquad (G \text{ is homologically spherical})$$

- of course, $S(p) \geq T(p)$
- plot of empirical functions $S(p)$ (resp. $T(p)$) computed as fractions of all homologically spherical (resp. trivial) digraphs among all available samples
  - see that for $p \approx 0.5$, a random digraph is *homologically spherical* with probability nearly 1, and is *homologically trivial* with probability close to 0.9

1.5 Computational limitations

- for computation of homology groups and Betti numbers of digraphs we use the aforementioned program by Chao Chen
- successively computes $H_k(G)$ and $\beta_k(G)$ for $k = 1, 2, \ldots$ until the memory of the computer allows
- N_a: the largest rank of actually computable Betti numbers for a digraph G
- for randomly generated digraphs with $p = q$, we have found the following empirical formula for N_a:

$$N_e = \frac{a \ln\left(1 + \frac{b}{V}\right)}{\ln D}$$

  where $D = \frac{E}{V}$ and a, b are constants to be found experimentally depending on the computer
  - for a 168GB i7 laptop, we have $a = 3, b = 400$

- if $D > 3$, then usually
$$|N_a - N_e| \leq 1$$
- since we have:
$$E \leq \frac{1}{2}V(V-1)$$
it follows that $D \leq \frac{1}{2}(V-1)$ and $V > 2D$. Therefore,
$$N_e \leq \frac{a \ln\left(1 + \frac{b}{2D}\right)}{\ln D}$$
  - with some data, the condition above implies that $D \leq 6$
    * hence, if for a randomly generated digraph, $D > 6$, then the computation of $\dim_h G$ becomes unreliable
- randomly generated digraph with $V = 30, E = 267, D = 8.9, p = q = 0.3$
  - we have that $N_e = 4$ while $N_a = 3$ and the actually computed Betti numbers are 1, 0, 0, 0.
  - since $D = 8.9 \gg N_a$, no reliable conclusion about $\dim_h G$ can be made
- for such digraphs, we need either to use a more powerful computer or to improve the algorithm of the program

Problem 3.6: Compute β_k for all $k \leq 9$ for the above digraph.

2 Curvature of digraphs

2.1 Motivation

- Γ: a finite planar graph
- there is the following classical notion of *curvature*

Definition 2.1. The **combinatorial curvature** K_x at any vertex x of Γ:
$$K_x := 1 - \frac{\deg(x)}{2} + \sum_{f \ni x} \frac{1}{\deg(f)} \qquad (4.1)$$
where the sum is taken over all faces f containing x and $\deg(f)$ denotes the number of vertices of f.

- e.g., if all faces are triangles, then we obtain:
$$K_x = 1 - \frac{\deg(x)}{2} + \frac{\deg_{\triangle}(x)}{3}$$
where $\deg_{\triangle}(x)$ is the number of triangles having x as a vertex.

- in general: denoting by E, V, F as the number of edges (resp. vertices, faces) of Γ and observing that:

$$\sum_x \deg(x) = 2E \qquad \text{and}$$
$$\sum_x \sum_{f \ni x} \frac{1}{\deg(f)} = \sum_f \sum_{x \in f} \frac{1}{\deg(f)} = F$$

we obtain:

$$\sum_x K_x = V - E + F = \chi \qquad \text{i.e., the Euler characteristic}$$

- we try to realize this idea on digraphs: to *distribute* the Euler characteristic over all vertices, and hence, to obtain an analogue of **Gauss curvature** that satisfies **Gauss-Bonnet**.

2.2 Curvature operator

- $G = (V, E)$: finite digraph
- $\mathbb{K} = \mathbb{R}$

Goal: To generalize (4.1) to arbitrary digraphs, so that the faces in (4.1) should be replaced by the elements of a basis in Ω_p – the faces should be replaced by the elements of a basis in Ω_p, but the result should be *independent* of the choice of a basis.

- fix $p \geq 0$
- any function $f : V \longrightarrow \mathbb{R}$ on the vertices induces a linear operator

$$T_f : \mathcal{R}_p \longrightarrow \mathcal{R}_p$$
$$T_f e_{i_0 \dots i_p} := (f(i_0) + f(i_1) + \cdots + f(i_p)) e_{i_0 \dots i_p}$$

- e.g., for a constant function $f = \mathbb{1}$ on V, we have $T_{\mathbb{1}} e_{i_0 \dots i_p} = (p+1) e_{i_0 \dots i_p}$ and hence

$$T_{\mathbb{1}} \omega = (p+1)\omega \text{ for any } \omega \in \mathcal{R}_p \qquad (4.3)$$

- if $f = \mathbb{1}_x$ where $x \in V$, then

$$T_{\mathbb{1}_x} e_{i_0 \dots i_p} = m e_{i_0 \dots i_p} \qquad (4.4)$$

where m is the number of occurrences of x in $i_0, \dots, i_p$

- in $\mathcal{R}_p$, fix an inner product $(\cdot, \cdot)$
  - e.g., this can be a **natural inner product** when all regular elementary paths $e_{i_0 \dots i_p}$ form an orthonormal basis in $\mathcal{R}_p$

- let $\Pi_p : \mathcal{R}_p \to \Omega_p$ be the orthogonal projection onto Ω_p
- considering T_f as an operator from Ω_p to $\mathcal{R}_p$, we obtain the following operator in Ω_p:

$$T'_f := \Pi_p \circ T_f : \Omega_p \to \Omega_p$$

Definition 2.2. The **incidence of** f **and** Ω_p is

$$[f, \Omega_p] := \mathrm{trace} T'_f$$

Definition 2.3. For any $\omega = \sum \omega^{i_0 \dots i_p} e_{i_0 \dots i_p} \in \Omega_p$, the **incidence of** f **and** ω is:

$$[f, \omega] := (T_f \omega, \omega)$$

Lemma 2.4. *For any orthogonal basis* $\{\omega_k\}$ *in* Ω_p*, we have:*

$$[f, \Omega_p] = \sum_k \frac{[f, \omega_k]}{\|\omega_k\|^2}$$

Proof. It suffices to prove the statement for an orthonormal basis (i.e., $\|\omega_k\| = 1$ for all k).

By the definition of the trace,

$$\mathrm{trace} T'_f = \sum_k (T'_f \omega_k, \omega_k)$$

For any $\omega \in \Omega_p$, we have

$$\begin{aligned} (T'_f \omega, \omega) &= \left(\Pi_p T_f \omega, \omega\right) \\ &= \left(T_f \omega, \Pi_p \omega\right) \\ &= (T_f \omega, \omega) \\ \rightsquigarrow \quad (T'_f \omega, \omega) &= [f, \omega] \end{aligned}$$

so that the statement follows. □

Definition 2.5. For any $N \in \mathbb{N}$, the **curvature operator** $K^{(N)} : \mathbb{R}^V \to \mathbb{R}$ **of order** N is

$$K^{(N)}(f) := \sum_{p=0}^{N} \frac{(-1)^p}{p+1} [f, \Omega_p]$$

- if $\Omega_p = \{0\}$ for all $p > N$, we write $K_f^{(N)} = K_f$
- for $f = \mathbb{1}_x$ for $x \in V$, we write

$$[x, \Omega_p] := [\mathbb{1}_x, \Omega_p] \text{ and}$$
$$[x, \omega] := [\mathbb{1}_x, \omega]$$

- if $\{\omega_k\}$ is an orthogonal basis of Ω_p, then by the preceding lemma, we have:

$$[x, \Omega_p] = \sum_k \frac{[x, \omega_k]}{\|\omega_k\|^2}$$

- if the inner product is natural so that $\{e_{i_0\ldots i_p}\}$ is an orthonormal basis, then by (4.4), we have:

$$[x, e_{i_0\ldots i_p}] = m \text{ where } m \text{ is the number of occurrences of } x \text{ in } i_0, \ldots, i_p$$

- as an example:

$$\begin{aligned}
[a, e_{abca}] &= 2\\
[b, e_{abca}] &= 1\\
[d, e_{abca}] &= 0
\end{aligned}$$

- in this case, for $\omega = \sum \omega^{i_0\ldots i_p} e_{i_0\ldots i_p}$, we have

$$[x, \omega] = \sum_{i_0\ldots i_p \in V} (w^{i_0\ldots i_p})^2 [x, e_{i_0\ldots i_p}]$$

Definition 2.6. Let $N \in \mathbb{N}$. The **curvature of order N at a vertex x** is

$$\begin{aligned}
K_x^{(N)} &:= K^{(N)} \mathbb{1}_x\\
\rightsquigarrow \quad K_x^{(N)} &= \sum_{p=0}^{N} \frac{(-1)^p}{p+1} [x, \Omega_p]
\end{aligned}$$

Proposition 2.7. ***Gauss-Bonnet.***

For any choice of the inner product in $\mathcal{R}_p$ and for any N, we have

$$\sum_{x \in V} K_x^{(N)} =: K_{total}^{(N)} = \chi^{(N)} := \sum_{p=0}^{N} \dim \Omega_p$$

Proof. Since $\sum_{x\in V} \mathbb{1}_x = \mathbb{1}$, we obtain that

$$\begin{aligned}
K_{\text{total}}^{(N)} &= \sum_{x \in V} K_x^{(N)}\\
&= \sum_{x \in V} K^{(N)} \mathbb{1}_x\\
&= K^{(N)} \mathbb{1}\\
\rightsquigarrow \quad K_{\text{total}}^{(N)} &= \sum_{p=0}^{N} (-1)^p \frac{[\mathbb{1}, \Omega_p]}{p+1}
\end{aligned}$$

On the other hand, by (4.3),

$$\begin{aligned}
[\mathbb{1}, \omega] &= (T_{\mathbb{1}} \omega, \omega)\\
&= (p+1) \|\omega\|^2
\end{aligned}$$

If $\{\omega_k\}$ is an orthogonal basis in Ω_p, then by (4.5),

$$
\begin{aligned}
[\mathbb{1}, \Omega_p] &= \sum_k \frac{[\mathbb{1}, \omega_k]}{\|\omega_k\|^2} \\
&= (p+1) \dim \Omega_p
\end{aligned}
$$

This implies:

$$
\begin{aligned}
& K^{(N)}_{\text{total}} = \sum_{p=0}^{N} (-1)^p \dim \Omega_p \\
\rightsquigarrow \quad & K^{(N)}_{\text{total}} = \chi^{(N)}
\end{aligned}
$$

□