Graph Mining by Vertex Dismantlings

Bertrand Jouve

LISST - CNRS - Toulouse - France

24/02/2020

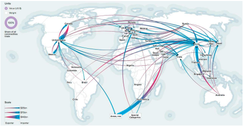
and Etienne Fieux

Plan

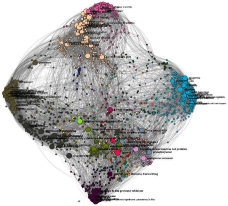
1. Introduction
   - A connected world
   - From local to global : two approaches
2. Model-based approach
   - Notations and definitions
   - Some well-known examples
   - About cycles in graphs ?
3. Clique complex
   - Simplicial and Strong Collapsibilities
   - Clique complex and k-dismantlability
   - k-homotopy
4. Conclusion

Understanding a connected world

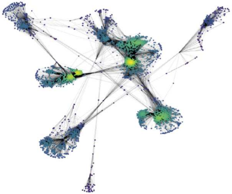
- The development of Big Data and data flow have produced a connected world made of complex interlocking networks.



resourcetrade.earth



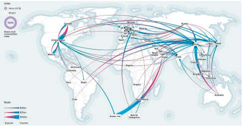
maps.gargantext.org



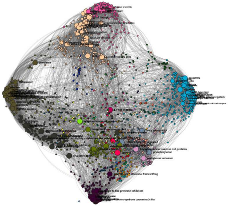
Social network

Understanding a connected world

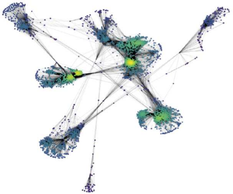
- The development of Big Data and data flow have produced a connected world made of complex interlocking networks.



resourcetrade.earth



maps.gargantext.org



Social network

- On these complex networks, an action at the global network level is often difficult and the global configuration is usually the result of the interactions of local (individual) actions.
- Understanding how these local interactions shape the network is necessary for understanding and possibly anticipating the development of these networks.
- If the network has special global properties then some computations are easier.

Understanding a connected world

- We focus on dynamics of network, i.e. how the shape of the network grows when new nodes are successively added ?
- Two types of approaches :

Data-based : For a real world network, learn rules from successive observations of new nodes in the network and then make predictions.

Model-based : From a single node, successively add new nodes with constraints for the attachment, and explore the global properties of the resulting network.

Understanding a connected world

- We focus on dynamics of network, i.e. how the shape of the network grows when new nodes are successively added ?
- Two types of approaches :

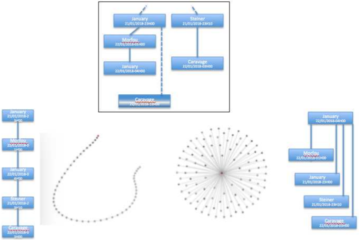
Data-based : For a real world network, learn rules from successive observations of new nodes in the network and then make predictions.

An example of data-based approach

When a lot of data are available, Machine Learning techniques can be used.

Example : Modeling discussion forums (Lumbreras et al., 2017)

Is it possible to replicate the topology of a discussion thread by observing where new messages appear ?

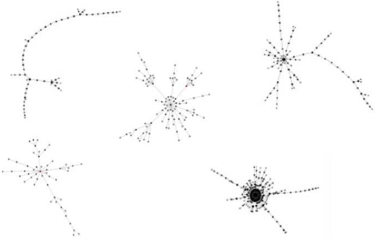


An example of data-based approach

When a lot of data are available, Machine Learning techniques can be used.

Example : Modeling discussion forums (Lumbreras et al., 2017)

Is it possible to replicate the topology of a discussion thread by observing where new messages appear ?



For a type *i* of users,

$$Attractivity(post) = \alpha_i \times popularity^{a_i}(post) + \beta_i \times \delta(root, post) + \tau_i \times date(post)$$

Understanding a connected world

- We focus on dynamics of network, i.e. how the shape of the network grows when new nodes are successively added ?
- Two types of approaches :

Model-based : From a single node, successively add new nodes with constraints for the attachment, and explore the global properties of the resulting network.

Notations and definitions

- G denote a finite undirected graph without loop, with n vertices and m edges.
- The open neighborhood of a vertex i in G : $N_G(i) = \{j; j \sim i\}$
- The closed neighborhood of a vertex i in G : $N_G[i] = N_G(i) \cup \{i\}$
- $G[1, \cdots, i]$ is the subgraph of G induced by vertices $1, \cdots, i$.

Definition

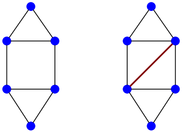
Given a property $\mathcal{P}$, a graph G has a vertex elimination scheme for $\mathcal{P}$ if there exists a linear ordering $1, 2, \cdots, n$ of its vertices such that $\forall i < n$, $N_{G-\{1,2,\cdots,i-1\}}(i)$ has property $\mathcal{P}$.

Remark : Equivalently, such graphs G can be constructed from an isolated vertex by successively adding vertices $1, 2, \cdots, n$ such that $\forall i > 1$, $N_{G[1,\cdots,i-1]}(i)$ has property $\mathcal{P}$.

Simpliciality ($\mathcal{P}$: " is complete")

Definition

A graph G is chordal if it contains no induced cycle of length ≥ 4



Simpliciality ($\mathcal{P}$: " is complete")

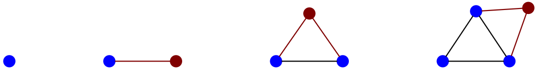
Definition

A graph G is chordal if it contains no induced cycle of length ≥ 4

Definition

- A vertex i is simplicial if $N_G(i)$ is complete.
- A graph G is simplicial if there is a linear ordering $1, 2, \cdots, n$ of its vertices such that $i < n$ is simplicial in $G - \{1, 2, \cdots, i - 1\}$.

Simplicial graphs can be constructed form the isolated vertex by successively adding vertices linked to a complete subgraph.



Simpliciality ($\mathcal{P}$: " is complete")

Definition

A graph G is chordal if it contains no induced cycle of length ≥ 4

Definition

- A vertex i is simplicial if $N_G(i)$ is complete.
- A graph G is simplicial if there is a linear ordering $1, 2, \cdots, n$ of its vertices such that $i < n$ is simplicial in $G - \{1, 2, \cdots, i-1\}$.

Simplicial graphs can be constructed form the isolated vertex by successively adding vertices linked to a complete subgraph.

Theorem (Dirac, 1961 ; Fulkerson and Gross, 1965)

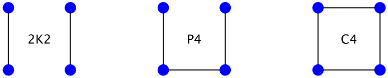
A finite graph is chordal if and only if it is simplicial.

Many NP-complete problems for general graphs are polynomial for chordal graphs (clique number, independence number, chromatic number, ...).

Threshold ($\mathcal{P}$: "is empty or the entire graph")

Definition

A threshold graph is a graph with no induced subgraph $2K_2$, P_4 or C_4.



Threshold ($\mathcal{P}$: "is empty or the entire graph")

Definition

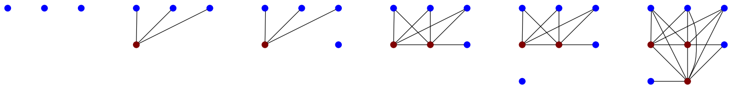
A threshold graph is a graph with no induced subgraph $2K_2$, P_4 or C_4.

Proposition (Chvátal & Hammer, 1975 ; Mahadev & Peled, 1995)

G is a threshold graph if, and only if, there is a linear ordering $1, \cdots, n$ of its vertices such that $\forall i < n$,

$$N_{G-\{1,\cdots,i-1\}}(i) = \emptyset \quad \text{or} \quad N_{G-\{1,\cdots,i-1\}}(i) = \{i+1, \cdots, n\}.$$

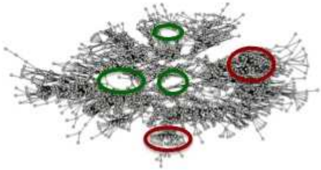
Threshold graphs can be constructed from the isolated vertex by successively adding either an isolated vertex or an universal vertex.



The importance of threshold graphs can be seen through numerous applications, from integer linear programming to partioning.

Structural holes in social networks

- Ronald. S. Burt (1982) “Structural holes : the social structure of competition” : Most sociologists agree that social structure is a kind of capital that can create for certain individuals or groups a competitive advantage in pursuing their ends. Burt introduced the concept of Structural Hole where an individual has to be to broker connections between otherwise “disconnected” segments.
- **We propose to focus on properties $\mathcal{P}$ related to cycles in graphs with the assumption that, with dense parts, they are major elements of the network structure.**



Generalizations of chordal graphs - Bridged graph ($\mathcal{P}$: "has a diameter $\leq$ 2")

Definition

- A chord of a cycle $\mathcal{C}$ is an edge joining 2 non-consecutive vertices of $\mathcal{C}$.
- A bridge of a cycle $\mathcal{C}$ is a shortest path joining 2 non-consecutive vertices of $\mathcal{C}$ and containing no edges of $\mathcal{C}$.
- A graph G is bridged if any cycle C of length ≥ 4 has a bridge.

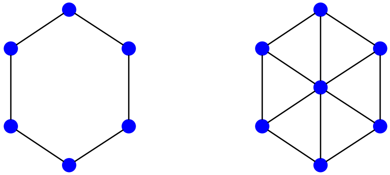


FIGURE: (a) Not bridged (b) Bridged but not Chordal

Generalizations of chordal graphs - Bridged graph ($\mathcal{P}$: "has a diameter ≤ 2")

Definition

- A graph G is bridged if any cycle C of length ≥ 4 has a bridge.

Definition

- A vertex i is isometric if the distances between the vertices of $G - i$ are equal to those between corresponding vertices in G.
- A graph G is isometric if there is a linear ordering $1, 2, \cdots, n$ of its vertices st. $i < n$ is isometric in $G - \{1, 2, \cdots, i - 1\}$.

Theorem (Anstee and Farber, 1988)

A finite graph is bridged iff it is isometric and has no induced C_4 or C_5.

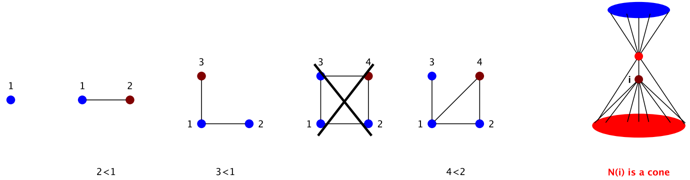
Isometric graphs can be constructed from the isolated vertex by successively adding vertices with neighborhoods of diameter ≤ 2 (ie. you must not shorten the distances).

Generalizations of chordal graphs - Dismantlable graph ($\mathcal{P}$: "is a cone")

Definition

- A vertex i is dominated or dismantlable in G if there exists $j \neq i$ such that $N_G[i] \subseteq N_G[j]$ (ie. $N_G[i]$ is a cone).
- A graph G is dismantlable if there is a linear ordering $1, 2, \cdots, n$ of its vertices st. $i < n$ is dismantlable in $G - \{1, 2, \cdots, i-1\}$.

Dismantlable graphs can be constructed from the isolated vertex by successively adding vertices whose neighborhoods are cones (convention : the isolated vertex is a cone).



On the Road to Algebraic Topology

- Even if G chordal $\Rightarrow$ G bridged $\Rightarrow$ G dismantlable, there is no easy characterization of a dismantlable graph in terms of cycles.
- But properties of cycles in bridged graphs (Jamison, 1985) and dismantlable graphs (Barmak & Minian, 2012 ; Fieux & Lacaze, 2012) may be expressed with concepts from algebraic topology.
- **The idea is to use concepts from algebraic topology to differentiate different types of networks according to the configurations of their induced cycles**. In particular, we translate the concept of homotopy (formalization of continuous deformation).

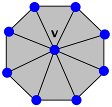
is not homotopic to
(you have to tear the ball)

Collapsibility of a simplicial complex

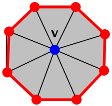
We need some notations and definitions.

- A simplicial complex $\mathcal{K}$, with vertex set V, is a collection of finite non empty subsets σ of V (the simplices) s.t. : $V = \bigcup_{\sigma \in \mathcal{K}} \sigma$ and if $(\sigma \in \mathcal{K}, x \in \sigma)$ then $\sigma - \{x\} \in \mathcal{K}$.
- A simplicial complex $\mathcal{K}$ is a cone with apex $a \in V$ if $a \cup \sigma \in \mathcal{K}$ for any simplex σ not containing a.
- For any vertex v of $\mathcal{K}$, $\mathrm{link}_{\mathcal{K}}(v) = \{\tau \in \mathcal{K} - v\ , \{v\} \cup \tau \in \mathcal{K}\}$.

Note that the link *is one of the possible notions to formalize an open neighborhood in case of simplicial complexes.*



A cone



link(v)

Collapsibility of a simplicial complex

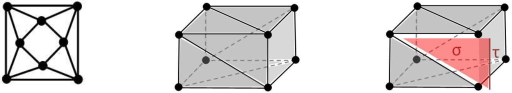
- An elementary strong collapse (Barmak & Minian, 2012) of $\mathcal{K}$ is the deletion of a vertex v whose *link* is a cone (ie. dominated vertex) and all the simplices containing v : $\mathcal{K} \searrow\searrow \mathcal{K} - \{v\}$.

Note that v is dominated in the 1-squelette of $\mathcal{K}$ considered as a graph.

Collapsibility of a simplicial complex

- An elementary strong collapse (Barmak & Minian, 2012) of $\mathcal{K}$ is the deletion of a vertex v whose *link* is a cone (ie. dominated vertex) and all the simplices containing v : $\mathcal{K} \searrow\searrow \mathcal{K} - \{v\}$.
- An elementary simplicial collapse (Whitehead, 1939) in $\mathcal{K}$ is the deletion of a pair of simplices (σ, τ) with τ a proper maximal face of σ and τ is not a face of another simplex : $\mathcal{K} \searrow \mathcal{K} - \{\sigma, \tau\}$.

Simplicial collapses are important transformations of simplicial complexes because they do not change the Betti numbers themselves related to the vector of simplicial cycles.



Collapsibility of a simplicial complex

- An elementary strong collapse (Barmak & Minian, 2012) of $\mathcal{K}$ is the deletion of a vertex v whose *link* is a cone (ie. dominated vertex) and all the simplices containing v : $\mathcal{K} \searrow\searrow \mathcal{K} - \{v\}$.
- An elementary simplicial collapse (Whitehead, 1939) in $\mathcal{K}$ is the deletion of a pair of simplices (σ, τ) with τ a proper maximal face of σ and τ is not a face of another simplex : $\mathcal{K} \searrow \mathcal{K} - \{\sigma, \tau\}$.
- $\mathcal{K} \searrow\searrow pt$: $\mathcal{K}$ is strong collapsible if it exists a sequence of elementary strong collapses such that $\mathcal{K} \searrow\searrow \cdots \searrow\searrow pt$.
- $\mathcal{K} \searrow pt$: $\mathcal{K}$ is collapsible if it exist a sequence of elementary simplicial collapses such that $\mathcal{K} \searrow \cdots \searrow pt$.

Proposition (Barmak & Minian, 2012)

Strong collapse is a particular kind of simplicial collapse :

$$\mathcal{K} \searrow\searrow pt \Rightarrow \mathcal{K} \searrow pt.$$

Collapsibility of a simplicial complex

Barmak & Minian note that it is possible to explore the gap between strong and simplicial collapses defining a k-collapsibility that, by induction, generalizes the strong collapsibility :

Definition (Barmak & Minian, 2012)

- $\mathcal{K} \searrow\!\!\!\searrow^{0} pt$: the 0-collapsability is the strong collapsability
- $\mathcal{K} \searrow\!\!\!\searrow^{n} \mathcal{K} - \{v\}$: $\mathcal{K}$ is n-collapsible to $\mathcal{K} - \{v\}$ if $\mathrm{link}_{\mathcal{K}}(v)$ is $(n-1)$- collapsible.
- $\mathcal{K} \searrow\!\!\!\searrow^{n} pt$: $\mathcal{K}$ is n-collapsible if it exists a sequence of n-collapses such that $\mathcal{K} \searrow\!\!\!\searrow^{n} \cdots \searrow\!\!\!\searrow^{n} pt$.

Proposition

$\mathcal{K} \searrow\!\!\!\searrow pt \Leftrightarrow \mathcal{K} \searrow\!\!\!\searrow^{0} pt \Rightarrow \mathcal{K} \searrow\!\!\!\searrow^{1} pt \Rightarrow \cdots \Rightarrow \mathcal{K} \searrow\!\!\!\searrow^{n} pt \Rightarrow \mathcal{K} \searrow pt.$

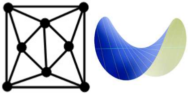
In practice, the presence of a cycle of length ≥ 4 often appears as a barrier to a dismantlabiity process and the smallest simplicial complexe on which a complex may be n-dismantlable is, in some ways, a composition of the remaining cycles. In Boulet, Fieux, J. (2008, 2010) and Fieux, J. (submitted) we show that this approach can be translated for graphs.

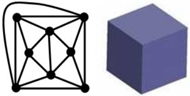
Clique complex and k-dismantlability

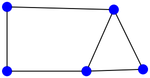
Given a simple graph G, consider the following complex :

Definition (Clique complex)

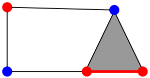
The clique complex $\Delta(G)$ of G is the simplicial complex such that $V(\Delta(G)) = V(G)$ and whose simplices are the cliques of G.



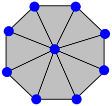




G



Δ(G) and link(1) in red



A cone

Clique complex and k-dismantlability

Given a simple graph G, consider the following complex :

Definition (Clique complex)

The clique complex $\Delta(G)$ of G is the simplicial complex such that $V(\Delta(G)) = V(G)$ and whose simplices are the cliques of G.

Now, define on graphs the concept of k-dismantlability which is a translation of the k-collapsibility :

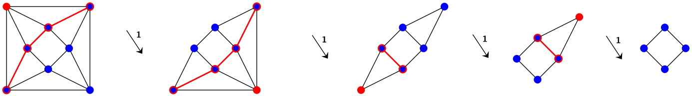
Definition (Boulet Fieux J., 2008, 2010, submitted)

- A vertex i is 0-dismantlable if it is dismantlable.
- A vertex i is 1-dismantlable if $N_G(i)$ is dismantlable and a graph G is 1-dismantlable if there is an ordering $1, 2, \cdots, n$ of its vertices such that $i < n$ is 1-dismantlable in $G - \{1, 2, \cdots, i-1\}$
- A vertex i is k-dismantlable if $N_G(i)$ is $(k-1)$-dismantlable. A graph G is k-dismantlable if there is an ordering $1, 2, \cdots, n$ of its vertices such that $i < n$ is k-dismantlable in $G - \{1, 2, \cdots, i-1\}$

Clique complex and k-dismantlability

Theorem

- *G is 0-dismantlable $\Leftrightarrow$ $\Delta(G)$ is 0-collapsible.*
- *G is k-dismantlable $\Leftrightarrow$ $\Delta(G)$ is k-collapsible.*
  More generally, $G \searrow^{k} H \Leftrightarrow \Delta(G) \searrow^{k} \Delta(H)$.

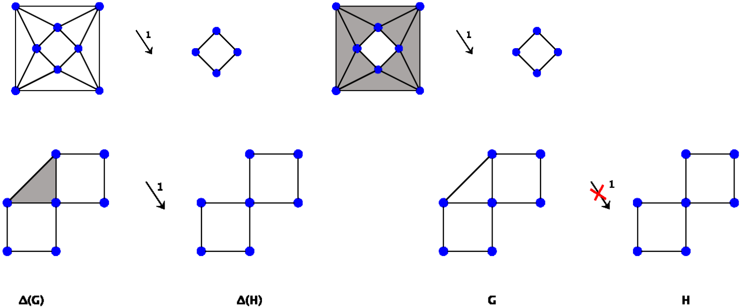


Clique complex and k-dismantlability

Theorem

- *G is 0-dismantlable* $\Leftrightarrow \Delta(G)$ *is 0-collapsible.*
- *G is k-dismantlable* $\Leftrightarrow \Delta(G)$ *is k-collapsible.*
  More generally, $G \searrow^{k} H \Leftrightarrow \Delta(G) \searrow^{k} \Delta(H)$.

Remark 1 : $G \searrow^{k} H \Rightarrow \Delta(G) \searrow \Delta(H)$ but the reverse implication is false.



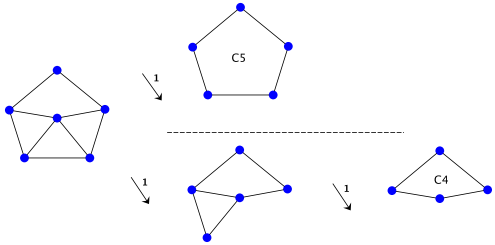
Clique complex and k-dismantlability

Theorem

- *G is 0-dismantlable* $\Leftrightarrow \Delta(G)$ *is 0-collapsible.*
- *G is k-dismantlable* $\Leftrightarrow \Delta(G)$ *is k-collapsible.*
  More generally, $G \searrow\searrow^{k} H \Leftrightarrow \Delta(G) \searrow\searrow^{k} \Delta(H)$.

Remark 1 : $G \searrow\searrow^{k} H \Rightarrow \Delta(G) \searrow\searrow \Delta(H)$ but the reverse implication is false.

Remark 2 : the order of k-dismantlability is important for $k \geq 1$:



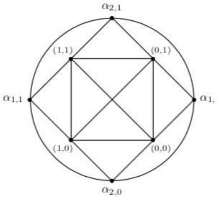
Clique complex and k-dismantlability

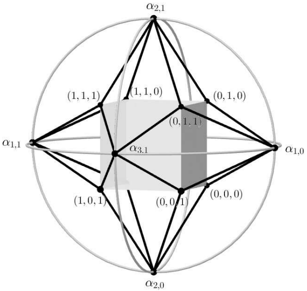
We denote by D_k the set of k-dismantlable graphs and D_{-1} the set of cones.

Proposition

The sequence $(D_k)_{k\geq 1}$ is strictly increasing :

$$D_{-1} \subsetneq D_0 \subsetneq D_1 \subsetneq D_2 \subsetneq \ldots\ldots \subsetneq D_k \subsetneq D_{k+1} \subsetneq \ldots\ldots$$





The n-cubion $\mathfrak{Q}_n$ is built from the n-hypercube and a n-octahedron $\overline{nK_2}$ so that each vertex of the octahedron is the apex of a cone whose base is a face of the cube.

Proposition

$\forall n \geq 2, \mathfrak{Q}_n \in D_{n-1} \setminus D_{n-2}$

Clique complex and k-dismantlability

We have some results on cliques (the existence of cliques with n vertices is not independent from the existence of induced cycles of a certain length)

Theorem

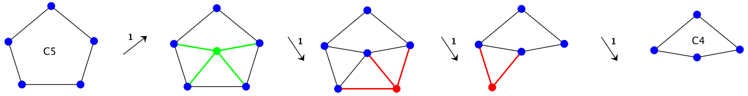
- *Given* $k \in \mathbb{N}$, *if* $G \in D_k \setminus D_{k-1}$, *then G contains a clique with at least* $k+2$ *vertices.*
- *Let G be a graph, if there is a clique A which intersects all maximal cliques of G, then either* $G \in D_{-1}$ *or* $G \in D_{a-2}$ *where* $a = |V(A)|$.

The second item refines the result of Rivest & Villemin (1976) proving that, given the same hypotheses, it exist k such that $G \in D_k$.

k-homotopy

The set D_k of k-dismantlable graphs is the set a graphs that can be obtained from the point by successively adding new vertices so that the neighborhood of each of these new vertices satisfies $(k-1)$-dismantlability. Let's get back to networks and assume now that we allow individuals to be added to or removed from the network.

- In terms of k-dismantlability, it means that we will now accept to both add and remove k-dismantlable vertices. In terms of simplicial complex, it means that we will now accept to both k-collapsibility and k-expansion.



- By translating the terms of algebraic geometry, we say that two graphs G and G' are k-homotopic if it is possible to go from G to G' by a succession of additons or deletions of k-dismantlble vertices. We write : $[G]_k = [G']_k$.

k-homotopy

We have essentially two results :

Proposition

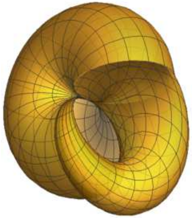
- $[G]_0 \neq [G]_1$ and for all integer $k \geq 1$, $[G]_1 = [G]_k$.
- $[G]_1 = [H]_1 \Leftrightarrow [\Delta(G)]_s = [\Delta(H)]_s$ where $[\cdot]_s$ is the simple simplicial homotopy (ie. add or remove free faces)

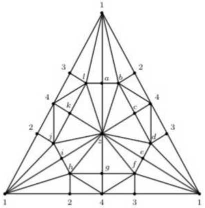
Remark (sensitivity)

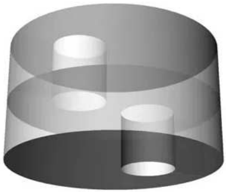
It is possible to find graphs that are not k-dismantlable for any k but which become 1-dismantlable by adding some 0-dismantlable vertices.

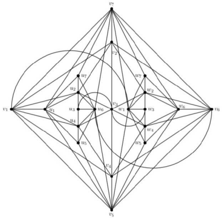
Two famous examples : The Dunce Hat and the Bing's House.

k-homotopy









The Dunce Hat, the Bing's House and the 1-skeletons of a triangulation. They are not collapsible but :

- After the 0-addition of 4 vertices the Dunce Hat becomes 1-dismantlable.
- After the 0-addition of 8 vertices the Bing's House becomes 1-dismantlable.

Conclusion

- Although we are often responsible for the complexity of large and dynamic networks, we don't fully understand or control its effects. If we collectively want to keep control of technology, this complexity must be understood and advanced mathematical tools may help.
- In complex networks, and in social network in particular, there exist many methods and efficient algorithms for studying the dense parts (communities) but little has been done to take into account "holes" in formalized dynamic models.
- Laurent & Tanigawa (2017) have extended to weighted graphs the notion of perfect elimination ordering, and so the notion of Chordal graphs. How about generalizing the k-dismantlability to weighted graphs ?
- A mathematical question : in topology continuous deformation of a simplicial complex to the point is the "contractibility" : k-collapsible and simplicial collapsible complex are contractible but the reverse is false. The Dunce Hat and the Bing's House are two counter-examples. Is it true that **if G is a graph such that $\Delta(G)$ is contractible, then it exist a graph W with $G \,{}_0\!\nearrow W \searrow_1 pt$?**

Thank you !

"Sorry, I cannot be with you today !"
(Etienne Fieux)