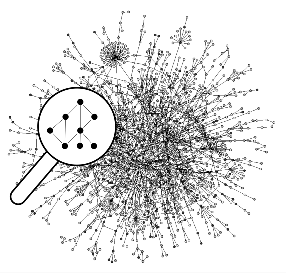
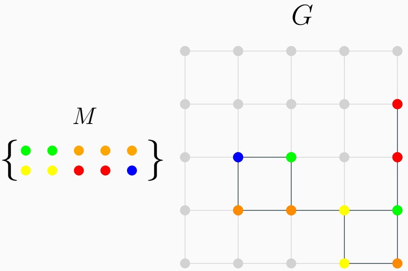
In the context of networks, the term motif may refer to different notions.

Subgraph motifs



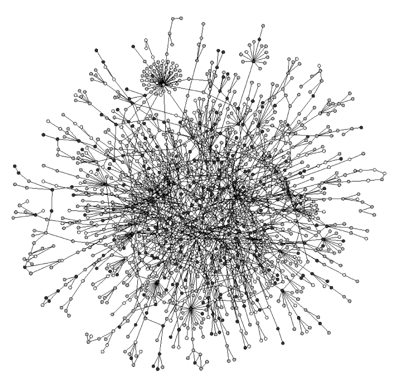
Coloured motifs



subgraph motifs

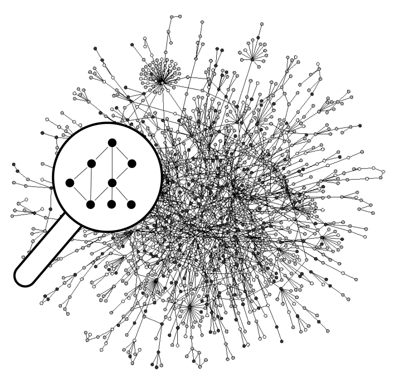
motifs

Find "interesting" patterns in a network.



motifs

Find "interesting" patterns in a network.



definition

Definition (Graph isomorphism)

Two graphs $G = (V, E)$ and $G' = (V', E')$ are said to be *isomorphic* if there exists a bijection $f : V \to V'$ such that for all u, v in V, $uv \in E \iff f(u)f(v) \in E'$.

definition

Definition (Graph isomorphism)

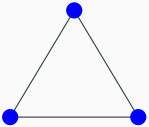
Two graphs $G = (V, E)$ and $G' = (V', E')$ are said to be *isomorphic* if there exists a bijection $f : V \to V'$ such that for all u, v in V, $uv \in E \iff f(u)f(v) \in E'$.

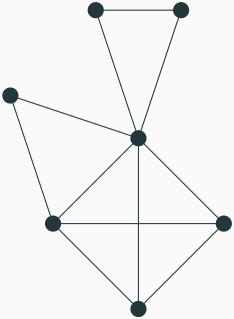
Let G, H be two graphs with $|V(H)| \leq |V(G)|$. An *occurrence* of H in G is a subset V' of vertices of G such that H and $G[V']$ are isomorphic.

definition

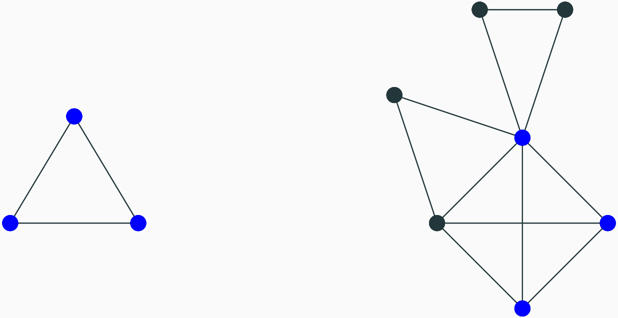
- If a graph H has at least an occurrence in a graph G, we say that G admit H as (induced) subgraph.
- We denote by $occ_G(H)$ the set of occurrences of H in G.
- The cardinality of $occ_G(H)$ is called the frequency of H in G

example

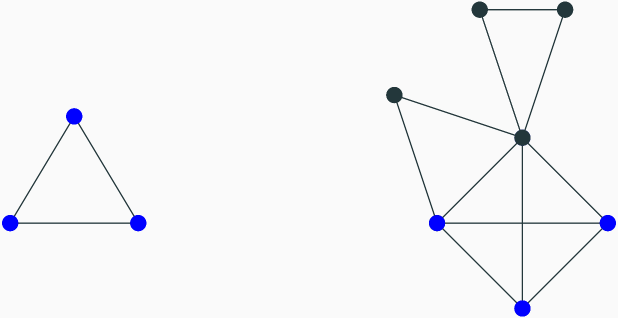




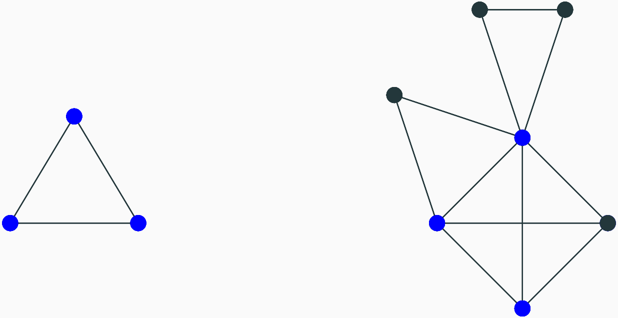
example



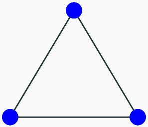
example

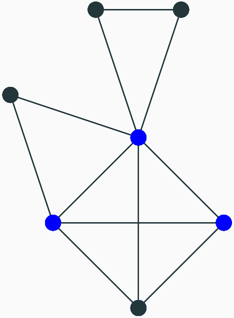


example

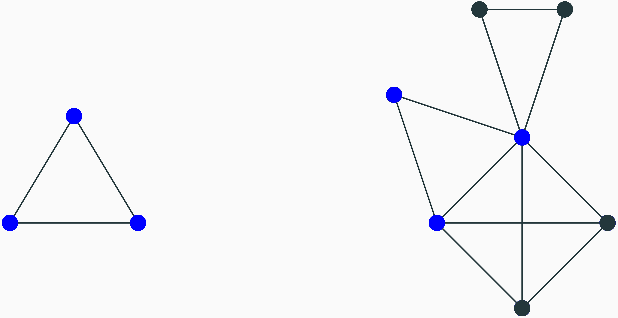


example

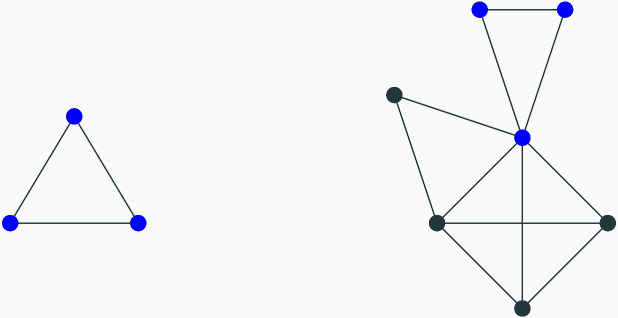




example



example



motifs

- Given a network, most of the time, some subgraphs are "overrepresented".
- A connected graph that has many occurrences in a network is called a *motif* of the network.

frequency

When can we say that a graph H is frequent in G?

frequency

When can we say that a graph H is frequent in G?

Naïve way :
Define a threshold. All graphs that have a frequency larger than the threshold are called frequent.

frequency

When can we say that a graph H is frequent in G?

Naïve way :

Define a threshold. All graphs that have a frequency larger than the threshold are called frequent.

The threshold usually depends on the size of H and G.

frequency

When can we say that a graph H is frequent in G?

Alternative way :
Compute the probability that $occ_N(H) \geq occ_G(H)$ for a random network N.

frequency

When can we say that a graph H is frequent in G?

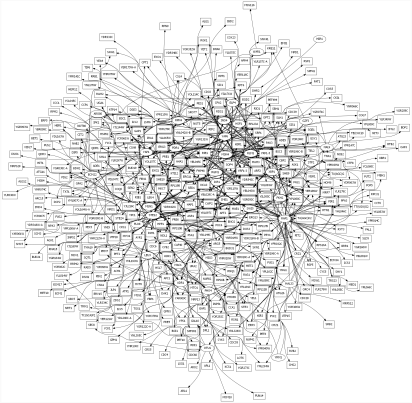
Alternative way :
Compute the probability that $occ_N(H) \geq occ_G(H)$ for a random network N.
H is said to be frequent in G is this probability is small enough.

When can we say that a graph H is frequent in G?

Alternative way :
Compute the probability that $occ_N(H) \geq occ_G(H)$ for a random network N.
H is said to be frequent in G is this probability is small enough.
To compute this probability, we need to have a distribution over networks.

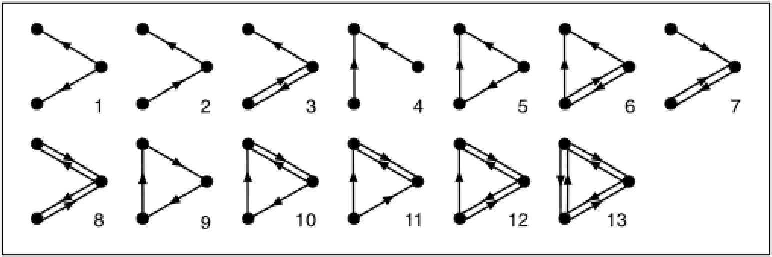
example: gene regulation network

A gene regulation network is an oriented graph. The vertices correspond to genes and there is an arc from g_1 to g_2 if the protein that g_1 encodes acts to alter the rate of expression of gene g_2.



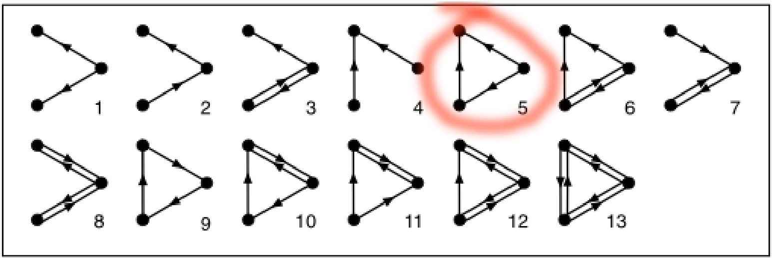
example: gene regulation network

Among all possible directed subgraphs of size three, one of them has a significant higher frequency than the others.



example: gene regulation network

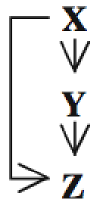
Among all possible directed subgraphs of size three, one of them has a significant higher frequency than the others.



It is called the “feed-forward loop”

feed-forward loop

The gene X regulates Z by two different ways.

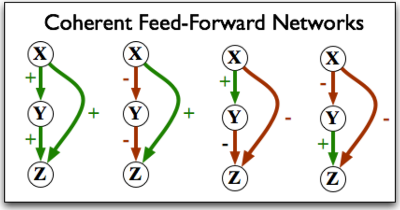


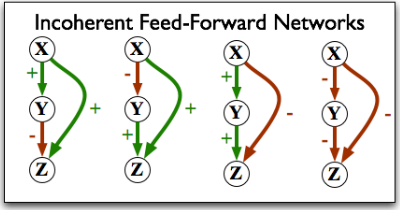
feed-forward loop

In a Gene regulation network we can label the arcs to precise if the $g1$ regulates g_2 positively or negatively.

feed-forward loop

In a Gene regulation network we can label the arcs to precise if the $g1$ regulates g_2 positively or negatively.
A feed-forward loop may correspond to several patterns





related problems

Problem: Subgraph isomorphism

Input: Two graphs H and G

Question Does H has at least one occurrence in G?

related problems

Problem: Subgraph isomorphism	
Input:	Two graphs H and G
Question	Does H has at least one occurrence in G?

The problem is NP-complete. Indeed, determining if a graph contains a clique of size k is already NP-complete

related problems

Problem: Subgraph isomorphism

Input: Two graphs H and G

Question Does H has at least one occurrence in G?

The problem is NP-complete. Indeed, determining if a graph contains a clique of size k is already NP-complete

Problem: Occurrences counting

Input: Two graphs H and G

Output Determine $occ_G(H)$.

This problem is $\#P$-complete.

Enumeration problems

Problem: Occurrences enumeration

Input: Two graphs H and G.

Output: $occ_G(H)$.

Enumeration problems

Problem: Occurrences enumeration	
Input:	Two graphs H and G.
Output:	$occ_G(H)$.

The problem is NP-hard since its associated decision problem is NP-complete.

Enumeration problems

Problem: Occurrences enumeration

Input: Two graphs H and G.

Output: $occ_G(H)$.

The problem is NP-hard since its associated decision problem is NP-complete.

Problem: Motifs enumeration

Input: A graph G.

Output The set of maximal frequent subraphs of G.

Enumeration problems

Problem: Occurrences enumeration

Input: Two graphs H and G.

Output: $occ_G(H)$.

The problem is NP-hard since its associated decision problem is NP-complete.

Problem: Motifs enumeration

Input: A graph G.

Output The set of maximal frequent subraphs of G.

The problem is NP-hard.

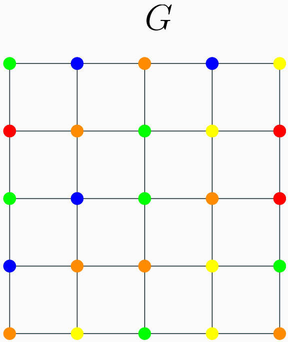
coloured motifs

coloured graph

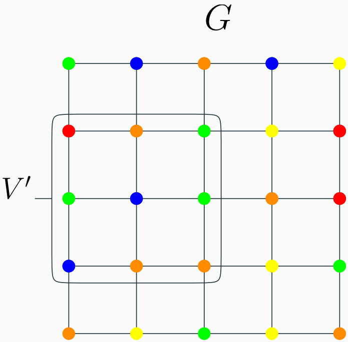
A colouration c of a graph $G = (V, E)$ is a map from V to a set of colours $\mathcal{C}$.

For a subset of vertices V' we denote by $col(V')$ the multiset of colours of V'.

example



example

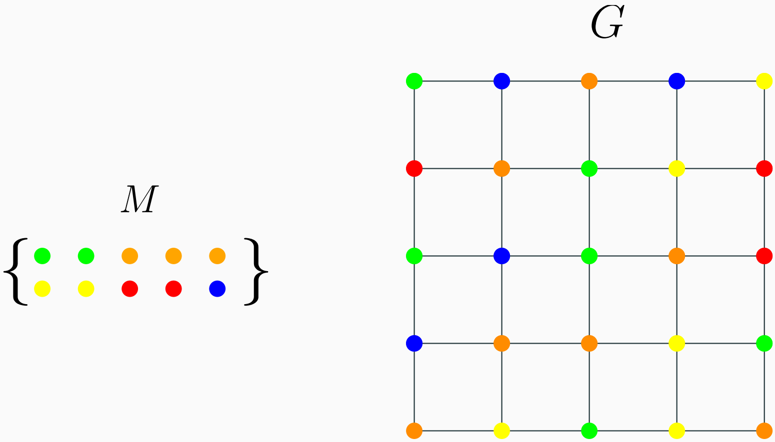


$col(V')$

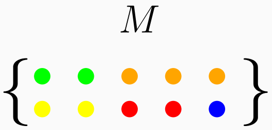
Coloured motif problem

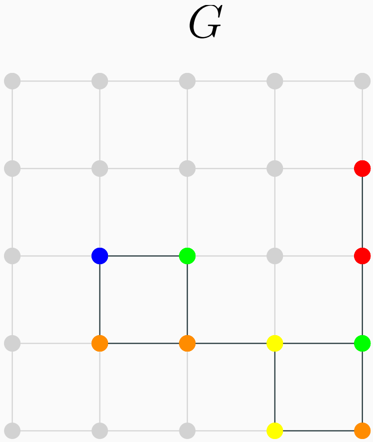
Input: A graph $G = (V, E)$ with a colouration $c : V \to \mathcal{C}$ and a multiset of colours M
Question: Is there a subset of vertices V' that induces a connected subgraph and such that $col(V') = M$?

example



example





colours

The colours model the similarity between vertices.

- In a protein-protein interaction network, two proteins have the same colours if they are homologous.
- In a metabolic network two reactions have the same colours if they use similar enzymes.

notations

Let k be the size of the motif M and let c be the number of colours in M.

Definition

The motif M is colourful if $k = c$ (each colour appear at most once in M)

difficulty

The problem is NP-complete even if:

- The graph is a tree and M is colourful.
- $c = 2$ and the graph is bipartite.
- M is colourful and the graph is of diameter two.

tractability

The problem become polynomial if the size of the motif k is constant.
More precisely, the problem parametrized by k is FPT!. There exists an algorithm of complexity $O(poly(n)f(k))$ where $poly$ is a polynomial and f is a function that depends only on k.

colourful motifs of bounded size

When the motif is colourful and the size of the motif is bounded, there is a dynamic programming algorithm of complexity $O(n2^{2k})$.

colourful motifs of bounded size

When the motif is colourful and the size of the motif is bounded, there is a dynamic programming algorithm of complexity $O(n2^{2k})$.
Main idea of the algorithm :

- Try to construct a tree containing all colours
- Given a vertex u, there exists a tree rooted in u containing all colours of M if there is a neighbour v of u and a set colours S such that there exists a tree rooted in u containing all colours of S and a tree rooted in v containing all colours of $M \setminus S$

dynamic programming algorithm

- Given a vertex u and a set of colours $S \subseteq M$, we say that $D(u, S)$ is True if there exists a tree rooted in u colourful on S.

dynamic programming algorithm

- Given a vertex u and a set of colours $S \subseteq M$, we say that $D(u, S)$ is True if there exists a tree rooted in u colourful on S.
- $D(u, S)$ is True if and only if there exists $S' \subsetneq S$ and a neighbour v of u such that $D(u, S') = True$ and $D(v, S \setminus S') = True$.

dynamic programming algorithm

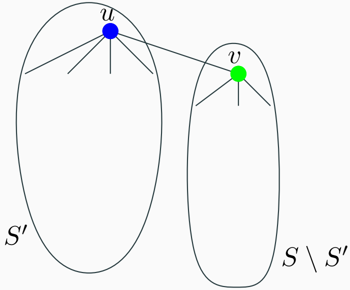
- Given a vertex u and a set of colours $S \subseteq M$, we say that $D(u, S)$ is True if there exists a tree rooted in u colourful on S.
- $D(u, S)$ is True if and only if there exists $S' \subsetneq S$ and a neighbour v of u such that $D(u, S') = True$ and $D(v, S \setminus S') = True$.
- Base cases: If $S = \{c_i\}$ then $D(u, S) = True$ iff $c_i = col(u)$

dynamic programming algorithm

- Given a vertex u and a set of colours $S \subseteq M$, we say that $D(u, S)$ is True if there exists a tree rooted in u colourful on S.
- $D(u, S)$ is True if and only if there exists $S' \subsetneq S$ and a neighbour v of u such that $D(u, S') = True$ and $D(v, S \setminus S') = True$.
- Base cases: If $S = \{c_i\}$ then $D(u, S) = True$ iff $c_i = col(u)$
- We build an $|V| \times 2^{|M|}$ $\mathcal{M}$ where the value of the cell $\mathcal{M}_{i,j}$ contains the value of $D(v, S)$ where v is the i^{th} vertex of the graph and S is the j^{th} subset of M.

dynamic programming algorithm

Decide if $D(u, S)$ is True.



dynamic programming algorithm

	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\cdots$	$\{2,3\}$	$\{1,2,3\}$
v_1							
v_2							
v_3							
$\vdots$ v_{n-1}	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
v_n							