From Path relations to phylogenetic trees

Yangjing Long

Ernst-Moritz-Arndt-Universität Greifswald
Center China Normal University
Joint work with **Volkmar Liebscher**

February 15, 2018

First theorem

Theorem

Yangjing was in Bled 6 years ago.

Proof.

Figure 1: 2012 2018

□

Question: Where is Greifswald?

Discussion Notes and Ideas

Dulce Valdivia[1], Alitzel López[1], Edgar Chávez[1], Marcos Gonzalez[1], Manuela Geiß[2], Peter F. Stadler[2], Marc Hellmuth[3], and Commandante Maribel Hernandez Rosales[1]

[1]The Mexican Math Cartel
[2]The Leipzig Connection
[3]Up There in the Wet, Cold North

November 14, 2017

Abstract

Where is Greifswald?

Figure 2: “Up There in the Wet, Cold North”== Greifswald

Motivation

- How much phylogenetic information can be extracted from single RGC (rare genomic changes)?

Motivation

- How much phylogenetic information can be extracted from single RGC (rare genomic changes)?
- How much information on the topology of the gene tree can be inferred from the knowledge of xenology relation?

Motivation

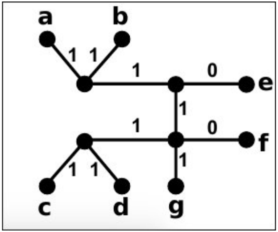
- How much phylogenetic information can be extracted from single RGC (rare genomic changes)?
- How much information on the topology of the gene tree can be inferred from the knowledge of xenology relation?
- These questions are to reconstruct phylogenetic trees from path properties(relations).

1. Path relations

2. How much Path relations are enough?

Edge labeled trees

Given (T, λ), a straightforward biological interpretation of an edge labeling $\lambda : E \to \{0, 1\}$ is that a certain type of evolutionary event has occurred along e if and only if $\lambda(e) = 1$. This suggests that in particular path properties and their associated relations on X are of practical interest:



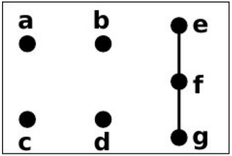


Figure 3: $x, y \in \Pi$ if and only if there is exactly one 1-edge along the path of x to y.

Edge labeled trees

Given (T, λ), a straightforward biological interpretation of an edge labeling $\lambda : E \to \{0, 1\}$ is that a certain type of evolutionary event has occurred along e if and only if $\lambda(e) = 1$. This suggests that in particular path properties and their associated relations on X are of practical interest:

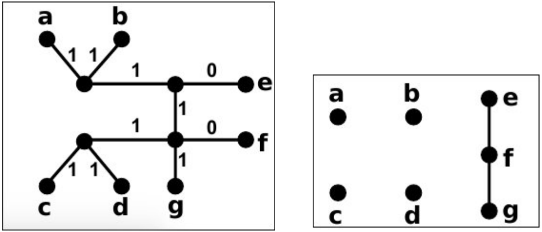


Figure 3: $x, y \in \Pi$ if and only if there is exactly one 1-edge along the path of x to y.

Question

How to reconstructe the tree and labelings from path relations?

Graph representation of $G(\overset{1}{\sim})/\overset{0}{\sim}$

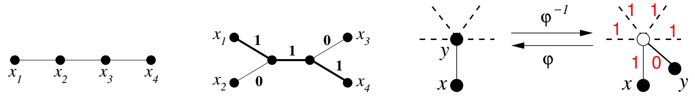
Theorem (Hellmuth, Hernandez-Rosales, L., Stadler 2017)

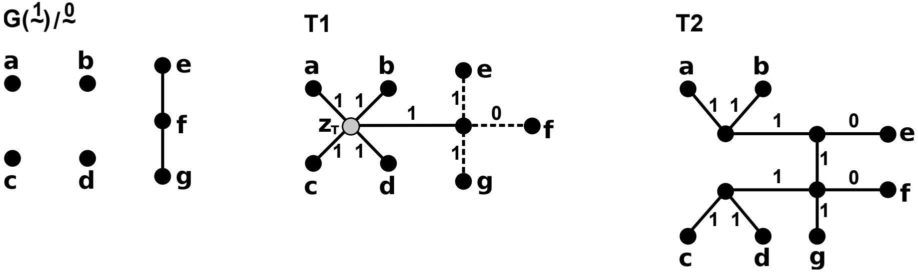
The graph $G(\overset{1}{\sim})/\overset{0}{\sim}$ is a forest.

Theorem (Hellmuth, Hernandez-Rosales, L., Stadler 2017)

If $G \in G(\overset{1}{\sim})/\overset{0}{\sim}$ is a tree, then the least resolved tree (T, λ) explains G is unique.

least resolved tree: no degree 2 vertices, no inner 0-edges.
We always consider least resolved tree through the talk!
Reconstruction





Theorem (Hellmuth, Hernandez-Rosales, L., Stadler 2017)

Let $Q_1, \ldots Q_k$ be the connected components in $G(\overset{1}{\sim})/\overset{0}{\sim}$. Up to the choice of the vertices q_i', the tree $T^* = T(G(\overset{1}{\sim})/\overset{0}{\sim}))$ is a minimally resolved tree that explains $G(\overset{1}{\sim})/\overset{0}{\sim}$. It is unique up to the choice of the $z_T q_i'$.

We also know something about the general setting

$x \overset{0}{\sim} y$ if and only if all edges in $\mathbb{P}(x, y)$ are labeled 0; For convenience we set $x \overset{0}{\sim} x$ for all $x \in X$.

$x \overset{1}{\sim} y$ if and only if all but one edges along $\mathbb{P}(x, y)$ are labeled 0 and exactly one edge is labeled 1;

$x \overset{1}{\rightharpoonup} y$ if and only if all edges along $\mathbb{P}(u, x)$ are labeled 0 and exactly one edge along $\mathbb{P}(u, y)$ is labeled 1, where $u = \text{lca}\{x, y\}$.

$x \overset{\geq k}{\sim} y$ with $k \geq 1$ if and only if at least k edges along $\mathbb{P}(x, y)$ are labeled 1;

$x \rightsquigarrow y$ if all edges along $\mathbb{P}(u, x)$ are labeled 0 and there are one or more edges along $\mathbb{P}(u, y)$ with a non-zero label, where $u = \text{lca}\{x, y\}$.

Our Main Question: Usually a path relation is not enough for tree reconstruction.

$\overset{1}{\sim}$ alone is not enough

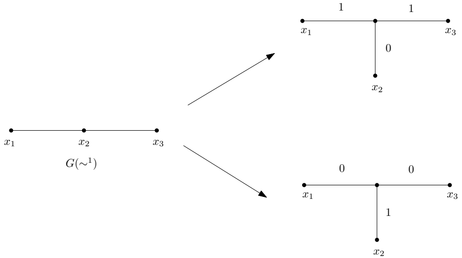


Figure 4: Two ways to reconstruct the tree from a connected $G(\overset{1}{\sim})$

Compare to tree metric, path relaiton has less information.

Volkmar Liebscher's question:

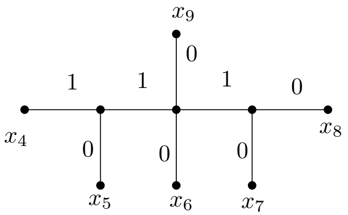
Question

When is single 1 relation enough to obtain the unique tree?
How much information is needed to add to the path relation to obtain tree metric?
How much information is needed to add to the path relation to obtain the unique tree with labelings?

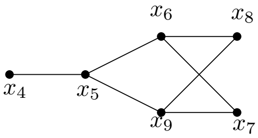
Sometimes $\overset{0}{\sim}$ and $\overset{1}{\sim}$ is enough

Lemma (Liebscher, L. 2018+)

$\overset{0}{\sim}$ and $\overset{1}{\sim}$ gives the tree metric if and only if $\overset{1}{\sim}$ is connected.



(T, λ)



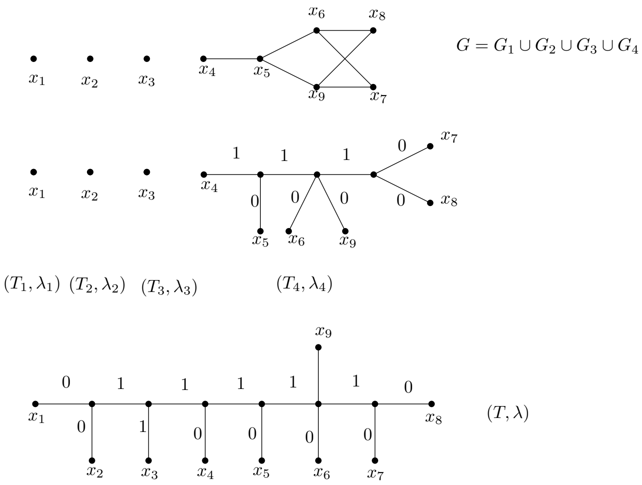


Figure 5: not enough information

Disconnected $\overset{1}{\sim}$

Proposition (Liebscher, L. 2018+)

$\overset{0}{\sim}$ *gives the tree metric if and only if* $G(\overset{0}{\sim})$ *is connected.*

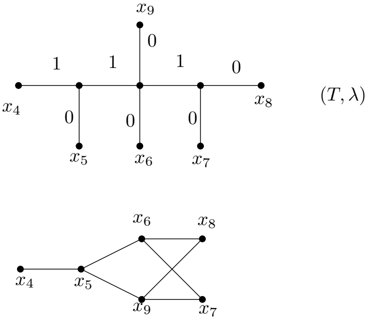
Proposition (Liebscher, L. 2018+)

$\overset{1}{\sim}$ *gives the tree metric if and only if* $G(\overset{1}{\sim})$ *is connected and point-determining.*

Connected $\overset{1}{\sim}$

Lemma (Liebscher, L. 2018+)

$\overset{0}{\sim}$ and $\overset{1}{\sim}$ give the tree metric if and only if $\overset{1}{\sim}$ is connected.



Theorem (Liebscher, L. 2018+)

The following statements are enough to characterize the tree (T, λ)*:*

(1) $\overset{0}{\sim}, \overset{1}{\sim}, \ldots, \overset{\geq k}{\sim}$ *are known where* k *is the smallest* k *such that* $\leq^k$ *is connected.*

(2) $\overset{0}{\sim}, \overset{1}{\sim}$ *and tree metric of quotient graph and informations of entrances are known.*

(3) $\overset{0}{\sim}, \overset{1}{\sim}$ *and* $\overset{\geq k}{\sim}$ *for all* k *where* k *is a distance of the clusters.*

(4) *all the path relation between any two leaves of the tree.*

Acknowledgement

- Thank *Peter* for always being supportive for last 10 years.
- Thank *Marc, Manuela, Maribel (Mareike)* etc.
- Thank *Yann* for getting me Wuhan connections, and for the wonderful ski trip!
- Thank *the organisers* for providing me the perfect time slot for my talk.


2018
新年快樂
HAPPY NEW YEAR