Tissue Transplantation Experiments: Inference and other Problems

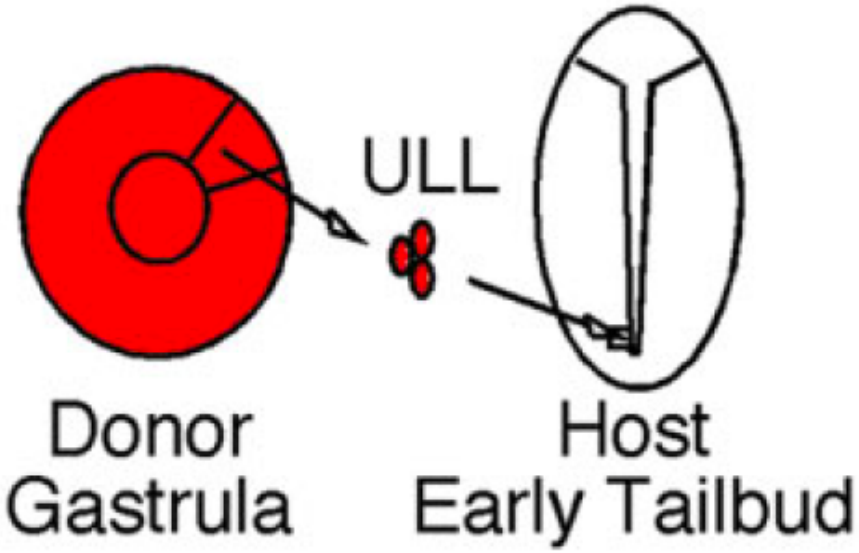
Yue Wang

Institut des Hautes Études Scientifiques (IHÉS), France

Dec. 17, 2020

Experiments

- Tissue transplantation experiments.



- Excise a piece of one tissue (donor), and transplant it to another tissue (host).

Experiments

The transplanted tissue is placed in an unfamiliar environment. Therefore it has many possible fates during development:

- Keep its original fate as donor tissue;
- Be assimilated by the host tissue;
- Transdifferentiate into a neither-donor-nor-host tissue.
- The development might be normal or abnormal.

Experiments

- Developmental biology: why could a zygote (in correct environment) develop into an adult animal.
- To understand why the developmental process in normal environment works, we also need to understand why the developmental process in abnormal environment does not work.
- Tissue transplantation experiments describe how tissues behave in abnormal environments.

Experiments

		Donor						
		AM19	PM19	PM15	UL11	LL11	LL15	LL19
	AM19	?	N	A	A	A	A	N
	PM19	?	N	?	N	N	?	?
H	PM15	?	?	?	?	?	?	?
o	UL11	?	N	?	N	N	?	?
s	LL11	?	N	?	N	N	?	?
t	LL15	?	?	?	?	?	?	?
	LL19	?	?	?	?	?	?	?

Table: Results reported by Krneta-Stankic et al. 2010

Only consider whether the result is normal (N) or abnormal (A).
AM19, PM19, etc.: different tissues.

Experiments

- There are not many known results of tissue transplantation experiments.
- We need a method to infer the unknown experimental results.
- We can use the known results and some biological common sense in the inference.
- First we need to discuss the similarities between tissues and between experiments.

Ideas

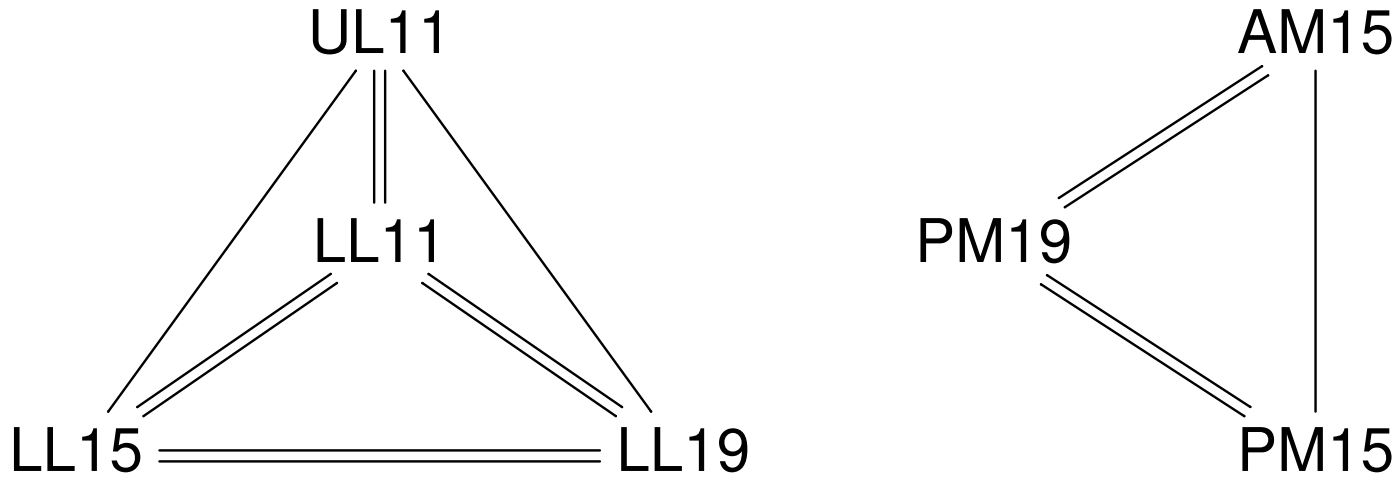
- We can describe the similarities between tissues qualitatively or quantitatively. (Transcriptome information, concentration of certain molecules, distance on the developmental tree, etc.)
- With the similarities between tissues, we can establish the similarities between experiments. Experiments are similar if they have similar donors and similar hosts.

Ideas

- Core idea 1: if donor and host tissues are (not) similar, the result tends to be (ab)normal.
- Core idea 2: similar experiments should have similar results.

Ideas

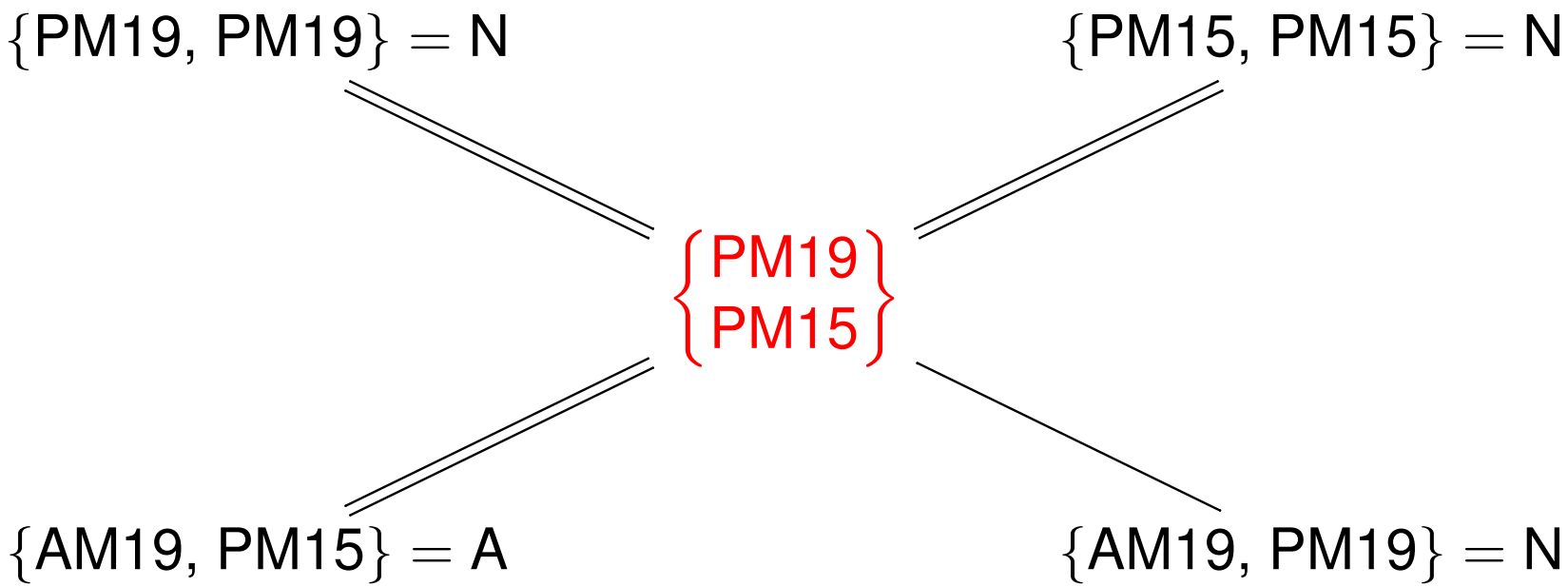
- Tissue similarity chart:



- Double/single/no line: high/medium/low similarity.
- For each experiment, we can make a prediction: if the donor and the host have high or medium similarity (e.g., {PM19, PM15}), the result tends to be normal; otherwise (e.g., {PM19, LL11}), the result tends to be abnormal.

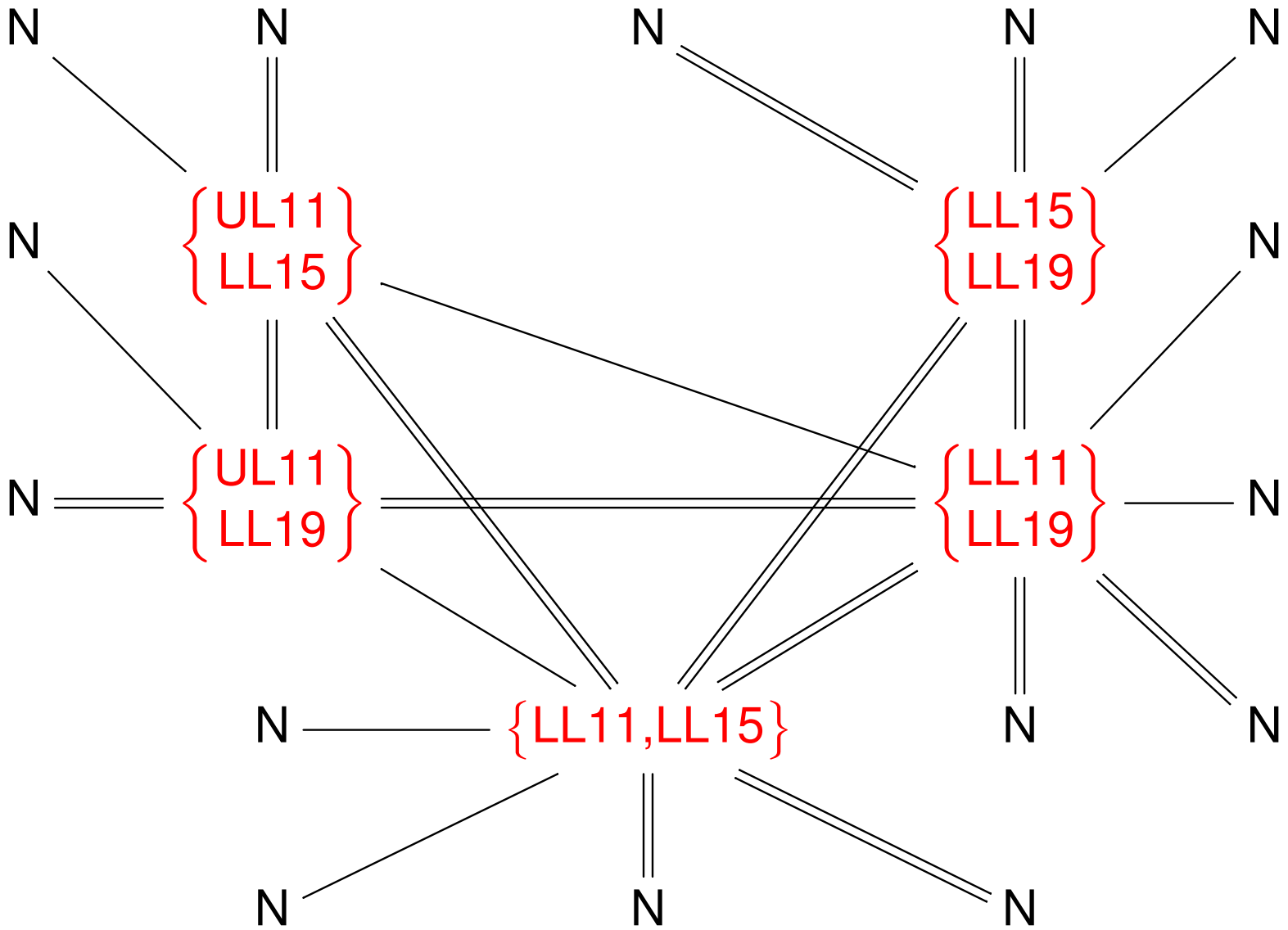
Ideas

Experiment similarity chart is determined by tissue similarity chart.
Black/red terms are experiments with known/unknown results.

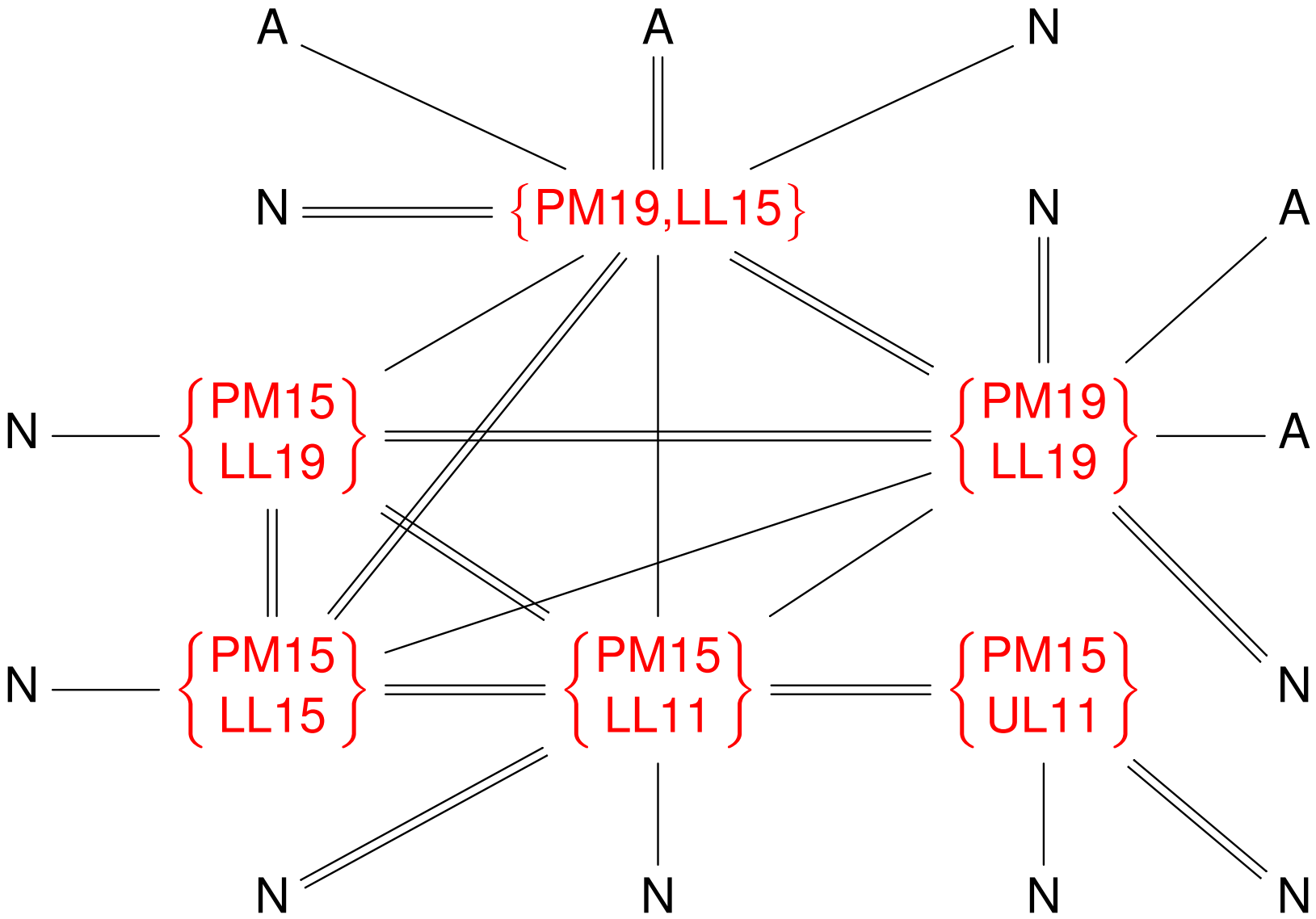


Double/single/no line: high/medium/low similarity.

Ideas


N
N
N
N
N
N
N
N
N
N
N
N
N
N
N
N
N
N
{UL11
LL15}
{LL15
LL19}
{UL11
LL19}
{LL11
LL19}
{LL11,LL15}

Ideas

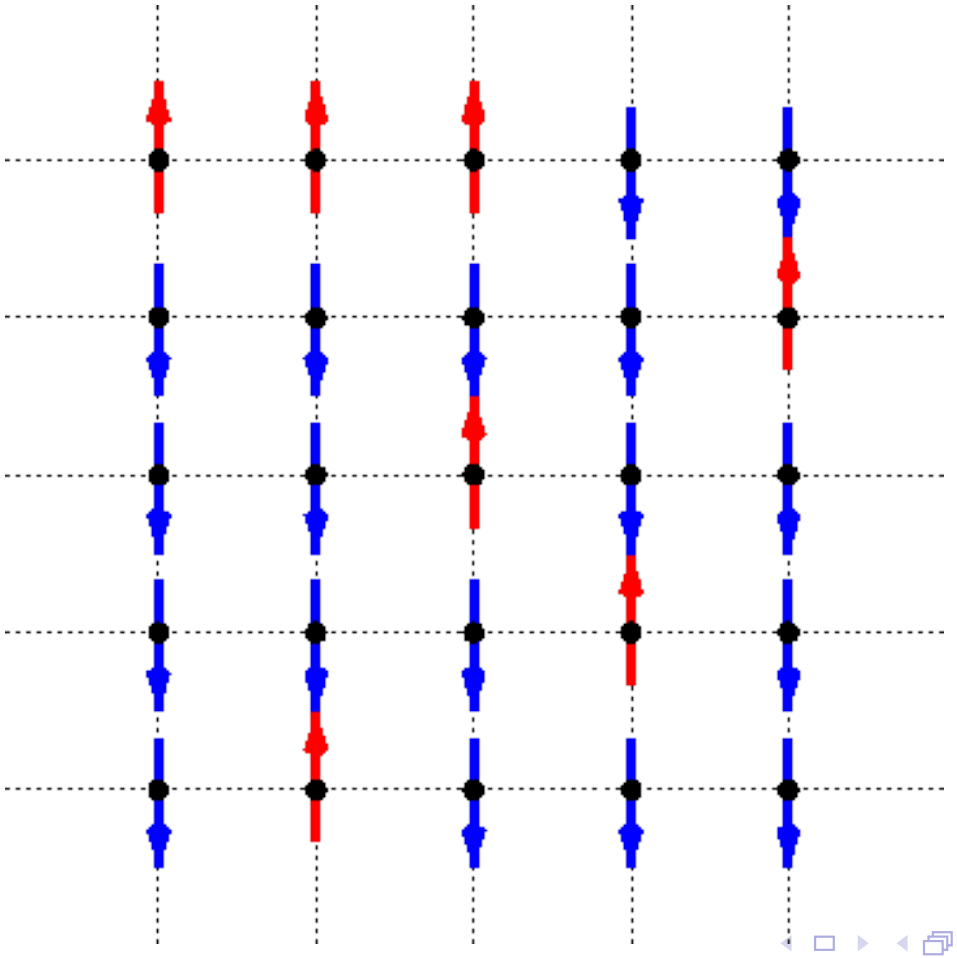

A
A
N
N
{PM19,LL15}
N
A
N
{PM15 LL19}
{PM19 LL19}
A
N
{PM15 LL15}
{PM15 LL11}
{PM15 UL11}
N
N
N
N
N

Ideas

- We take guesses for unknown experimental results, and design a penalty function to evaluate these guesses.
- There is a penalty if the result of an experiment violates our prediction.
- There is a penalty if two similar experiments have different results.
- For the concrete form of this penalty function, we can get inspirations from the Ising model.

Ising model

The Ising model describes ferromagnetism in statistical mechanics. Consider a set of lattice sites, where each site k has a variable σ_k that takes $+1$ or -1. Each site also has an external field.



Ising model

- For a configuration σ of ± 1, its energy function is

$$\mathrm{H}(\sigma) = -\sum_{i \sim j} J_{ij}\sigma_i\sigma_j - \sum_j h_j\sigma_j,$$

  where $i \sim j$ means site i and site j are neighboring, $J_{ij} \geq 0$ is the interaction coefficient, and h_j is the external field.
- The probability of a configuration σ is

$$\mathbb{P}_\beta(\sigma) = e^{-\beta \mathrm{H}(\sigma)}/Z_\beta,$$

  where $\beta = (k_B T)^{-1}$, Z_β is the normalization constant.
- Configuration with higher energy has smaller probability. The value tends to have the same sign with the external field. Neighboring sites tend to have the same value.

Ideas

Analogies between tissue transplantation experiments and the Ising model:

Tissue transplantation	Ising model
Experiment similarity chart	Lattice
Experiment	Site
Similar experiments	Neighboring sites
Result: normal/abnormal	Value: +1/-1
Prediction	External field
Penalty: result and prediction are different	Penalty: value and external field have opposite signs
Penalty: similar experiments have different results	Penalty: neighboring sites have different values
Penalty function?	Energy function

Pure analogy, not physical correspondence.

Ideas

- For tissue transplantation experiments, we take guesses for unknown experimental results $\{\sigma_i\}$.
- The penalty function is

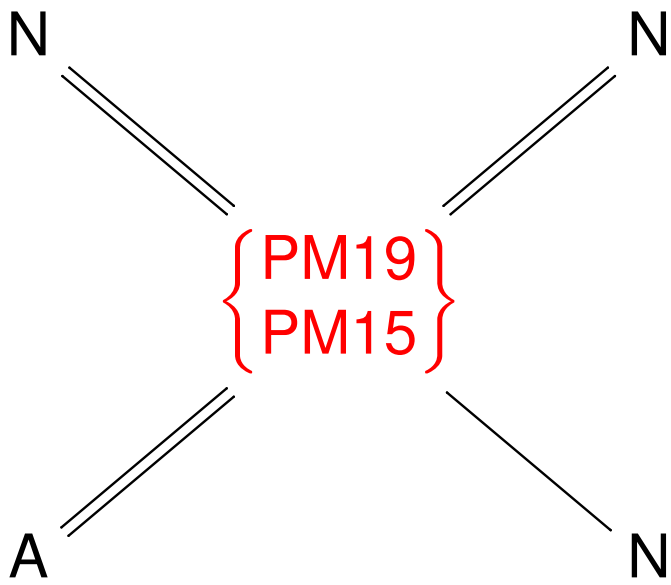
$$\mathrm{H}(\sigma) = -\sum_{i,j} J_{ij}\sigma_i\sigma_j - \sum_j h_j\sigma_j.$$

- The probability of a configuration σ is

$$\mathbb{P}_\beta(\sigma) = e^{-\beta \mathrm{H}(\sigma)}/Z_\beta.$$

- Regard N as $+1$, and A as -1.
- For two experiments σ_i, σ_j with high/medium/low similarity, set $J_{ij} = 2J_0/J_0/0$. Set $h_j = h_0\pi_j$, where π_j is prediction.
- The values of parameters J_0, h_0, β are yet to be determined by cross validation.

Ideas

- 

- Since PM19 and PM15 are highly similar, our prediction for experiment {PM19,PM15} is $\pi_j =$ N (normal).
- $\{\text{PM19}, \text{PM15}\} =$ N: $\mathrm{H} = -4$, $\mathbb{P} = 0.69$.
- $\{\text{PM19}, \text{PM15}\} =$ A: $\mathrm{H} = 4$, $\mathbb{P} = 0.31$.
- Result=N is the most probable guess. $\mathbb{P}(\text{N}) = 0.69$.

Ideas

Configuration					Penalty	Probability
-1	-1	-1	-1	-1	14	0.0019
1	-1	-1	-1	-1	10	0.0029
-1	1	-1	-1	-1	14	0.0019
1	1	-1	-1	-1	2	0.0064
-1	-1	1	-1	-1	14	0.0019
1	-1	1	-1	-1	2	0.0064
-1	1	1	-1	-1	6	0.0043
1	1	1	-1	-1	-14	0.0319
-1	-1	-1	1	-1	16	0.0016
1	-1	-1	1	-1	12	0.0024
-1	1	-1	1	-1	8	0.0035
1	1	-1	1	-1	-4	0.0117
-1	-1	1	1	-1	12	0.0024
1	-1	1	1	-1	0	0.0079
-1	1	1	1	-1	-4	0.0117
1	1	1	1	-1	-24	0.0868

Ideas

For each configuration of the unknown results (guesses), we can calculate its probability. We can determine the most probable guesses:

		Donor						
		AM19	PM19	PM15	UL11	LL11	LL15	LL19
	AM19	*N*	N	A	A	A	A	N
	PM19	*N*	N	N	N	N	N	N
H	PM15	*A*	N	*N*	N	N	N	N
o	UL11	*A*	N	N	N	N	N	N
s	LL11	*A*	N	N	N	N	N	N
t	LL15	*A*	N	N	N	N	*N*	N
	LL19	*N*	N	N	N	N	N	*N*

Ideas

We can also calculate the expectation of all guesses, i.e., the probability for each experimental result to be “N”:

		Donor						
		AM19	PM19	PM15	UL11	LL11	LL15	LL19
	AM19	*100%*	100%	0%	0%	0%	0%	100%
	PM19	*100%*	100%	65%	100%	100%	49%	56%
H	PM15	*0%*	65%	*100%*	62%	62%	53%	54%
o	UL11	*0%*	100%	62%	100%	100%	81%	81%
s	LL11	*0%*	100%	62%	100%	100%	90%	90%
t	LL15	*0%*	49%	53%	81%	90%	*100%*	86%
	LL19	*100%*	56%	54%	81%	90%	86%	*100%*

Another situation

- For now, we have designed an inference method that works for experiments with binary deterministic results.
- What if the known experimental results are not deterministic, but stochastic?

Another situation

		Donor				
		PLE11	PLE12	PLE14	PLE16	PLE19
	LFR\PLE14	61%	58%	82%	?	?
Host	LFR\PLE16	?	?	?	?	?
	LFR\PLE19	4%	24%	83%	?	100%

Table: Results reported by Henry et al. 1987

- Percentage is lens formation (normal growth) rate.
- PLE11: presumptive lens ectoderm, stage 11.
- LFR\PLE14: lens-forming region without presumptive lens ectoderm, stage 14.

Another situation

Sample deterministic configurations from these stochastic results (assume different experiments are independent). For each deterministic configuration, apply our method to obtain the expectation of guesses.
For example, assume we have three similar experiments: [61%N ? 58%N].

$$\mathbb{P}([\text{N ? N}]) = 61\% \times 58\% = 35\%.$$

$$\mathbb{P}(\text{?=N} \mid [\text{N ? N}]) = 98\%.$$

$$\mathbb{P}([\text{N N N}]) = \mathbb{P}([\text{N ? N}]) \times \mathbb{P}(\text{?=N} \mid [\text{N ? N}]) = 35\% \times 98\% = 35\%.$$

Another situation

Similarly, we can calculate for other deterministic configurations of [61%N ? 58%N]:

$$\mathbb{P}([\text{N ? A}]) = 61\% \times (100\% - 58\%) = 26\%.$$

$$\mathbb{P}([\text{N N A}]) = 26\% \times 50\% = 13\%.$$

$$\mathbb{P}([\text{A ? N}]) = (100\% - 61\%) \times 58\% = 23\%.$$

$$\mathbb{P}([\text{N N A}]) = 23\% \times 50\% = 11\%.$$

$$\mathbb{P}([\text{A ? A}]) = (100\% - 61\%) \times (100\% - 58\%) = 16\%.$$

$$\mathbb{P}([\text{A N A}]) = 16\% \times 2\% = 0\%.$$

Then average over these deterministic configurations:

$$\mathbb{P}([\text{?=N}]) = \mathbb{P}([\text{N N N}]) + \mathbb{P}([\text{N N A}]) + \mathbb{P}([\text{A N N}]) + \mathbb{P}([\text{A N A}]) = 59\%.$$

Final results: [61%N 59%N 58%N].

Another situation

				Donor		
		PLE11	PLE12	PLE14	PLE16	PLE19
	LFR\PLE14	61%	58%	82%	?	?
Host	LFR\PLE16	?	?	?	?	?
	LFR\PLE19	4%	24%	83%	?	100%

Table: Results reported by Henry et al. 1987

				Donor		
		PLE11	PLE12	PLE14	PLE16	PLE19
	LFR\PLE14	61%	58%	82%	93%	94%
Host	LFR\PLE16	39%	53%	88%	97%	97%
	LFR\PLE19	4%	24%	83%	96%	100%

Yet another situation

- The methods are for experiments with binary results.
- What if the known experimental results are not binary?
- Original penalty function:

$$\mathrm{H}(\sigma) = -\sum_{i \sim j} J_{ij}\sigma_i\sigma_j - \sum_j |h_j|\pi_j\sigma_j.$$

Yet another situation

- Rewrite the penalty function:

$$\mathrm{H}(\sigma) = -\sum_{i\sim j} J_{ij} f(\sigma_i, \sigma_j) - \sum_j |h_j| f(\pi_j, \sigma_j).$$

- If σ_i, σ_j are more similar, $f(\sigma_i, \sigma_j)$ is larger. Also, $f(\sigma_i, \sigma_j) = f(\sigma_j, \sigma_i)$. For binary case, $f(\sigma_i, \sigma_j) = \sigma_i \sigma_j$.
- Example: In the chick experiments by Hamburger (1939), there are three possible results, ND, AD, TA. We can define f as:

$$f(\text{ND,ND}) = 2;\ f(\text{ND,AD}) = 0;\ f(\text{ND,TA}) = -1;$$

$$f(\text{AD,ND}) = 0;\ f(\text{AD,AD}) = 2;\ f(\text{AD,TA}) = 0;$$

$$f(\text{TA,ND}) = -1;\ f(\text{TA,AD}) = 0;\ f(\text{TA,TA}) = 2.$$

Summary so far

- Based on the similarities between tissues and between experiments, we have designed methods to infer the unknown experimental results.
- The results are not necessarily deterministic or binary.
- Such methods should not be limited to tissue transplantation experiments.

- Questions?

Experimental design

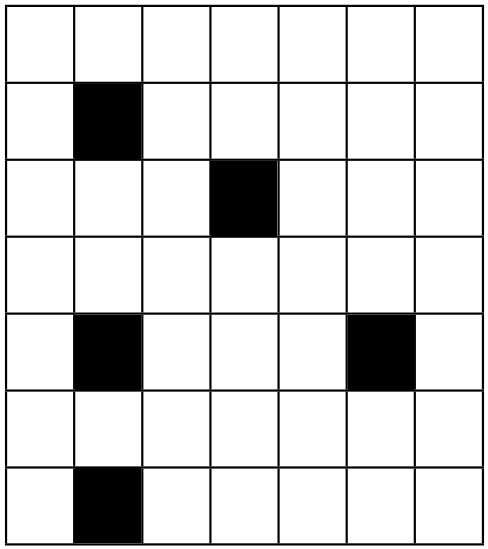
- There are many tissue transplantation experiments, and we want to know all the results.
- Assume we do not have any results yet. We can choose some experiments to conduct, and use the known results to infer other experiments.
- How to choose experiments to conduct?
- We need enough data to apply the inference method. We should minimize the experimental cost.

Experimental design

- The results of non-conducted experiments are inferred by similar conducted experiments.
- To guarantee the inference quality, one non-conducted experiment should be similar to at least k conducted experiments (e.g., $k = 2$ or $k = 1$).
- The most efficient design: no conducted experiments are similar, and each non-conducted experiment is similar to exactly k conducted experiments.

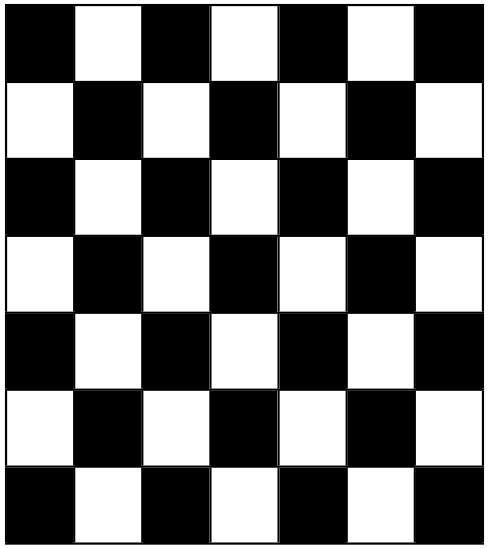
Experimental design

- Consider the figure that each unit is an experiment, and neighboring units are similar experiments. For now, assume the figure is $\mathbb{Z}^2$.
- Black units are conducted experiments, and white units are non-conducted experiments.
- How to color the figure, so that two black units are not neighboring, and each white unit is neighboring to k black units?



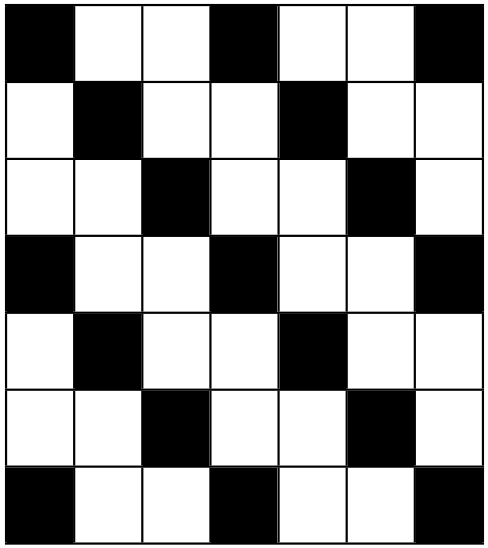
Experimental design

No neighboring black units, and each white unit is neighboring to $k = 4$ black units (ignore the boundary cases).



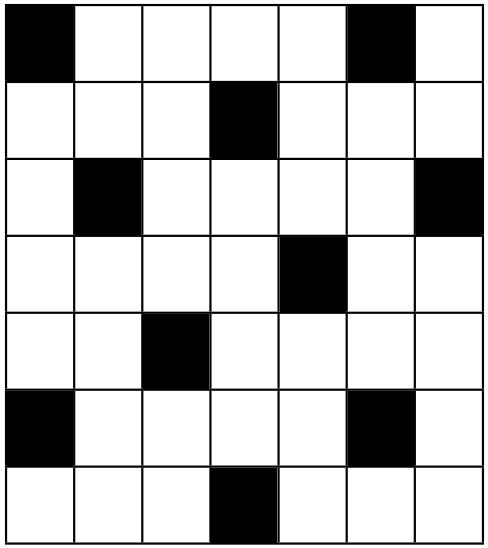
Experimental design

No neighboring black units, and each white unit is neighboring to $k = 2$ black units (ignore the boundary cases).



Experimental design

No neighboring black units, and each white unit is neighboring to $k = 1$ black unit (ignore the boundary cases).



Experimental design

- For $k = 4$, we need to conduct $1/2$ experiments.
- For $k = 2$, we need to conduct $1/3$ experiments.
- For $k = 1$, we need to conduct $1/5$ experiments.

Experimental design

- In practice, the experiment similarity chart is not 2-dimensional, but 4-dimensional. Each experiment has a coordinate (x, y, z, w) that stands for donor tissue type, donor tissue developmental stage, host tissue type, host tissue developmental stage.
- For now, we assume the chart is $\mathbb{Z}^4$.
- Similar coloring problems for such 4-dimensional figures.
- We need some abstract methods.

Experimental design

- In $\mathbb{Z}^4$, each unit is neighboring to 8 units.
- For $k = 8$, color a unit (x, y, z, w) if

$$x + y + z + w \equiv 0 \mod 2.$$

  We need to conduct $1/2$ experiments.
- For $k = 4$, color a unit (x, y, z, w) if

$$x + y + z + w \equiv 0 \mod 3.$$

  We need to conduct $1/3$ experiments.
- For $k = 2$, color a unit (x, y, z, w) if

$$x + 2y + z + 2w \equiv 0 \mod 5.$$

  We need to conduct $1/5$ experiments.

Experimental design

- For $k = 1$, color a unit (x, y, z, w) if

$$x + 2y + 3z + 4w \equiv 0 \mod 9.$$

  We need to conduct $1/9$ experiments.
- For this case, each white unit corresponds to 1 black unit, and each black unit corresponds to 8 white units. Thus the black-white ratio is $1 : 8$, and the proportion of black units is $1/9$. This is why we use $\mod 9$.
- It is easy to verify this construction works. However, it is difficult to explain how to obtain this construction. There might be deeper mathematics hidden behind.

Experimental design

- In practice, $k = 2$ or $k = 1$ is enough to conduct satisfactory inference. Therefore we only need to conduct $1/5 - 1/3$ experiments (two-dimensional) or $1/9 - 1/5$ experiments (four-dimensional).
- In practice, we might have more experimental similarities than $\mathbb{Z}^4$. Therefore we need to conduct fewer experiments. (If there are more similarity relationships, the same known result can provide more information about its unknown neighbors.)

Summary

- Penalty function-based methods to infer the unknown experimental results.
- Combinatorial experimental design that the inference methods can be applied most efficiently.

Reference

Yue Wang, Jérémie Kropp, and Nadya Morozova. (2020). “Inference on tissue transplantation experiments.” Preprint on arXiv: 2010.02704. Submitted.

Thank you!