

方言群の語彙は 系統樹をなすか

九州大学
村脇 有吾



本日のおはなし

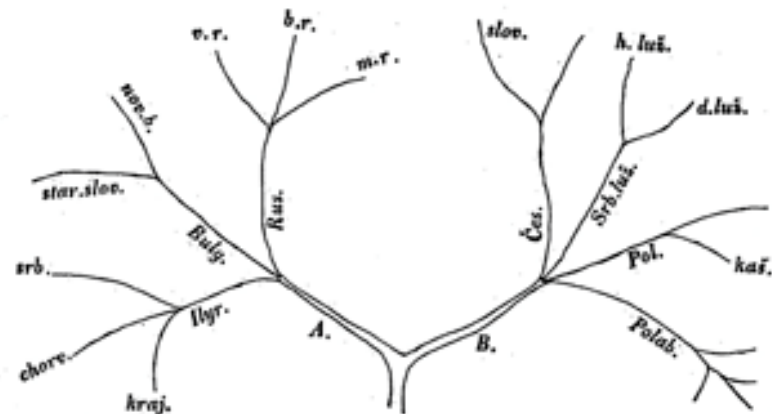
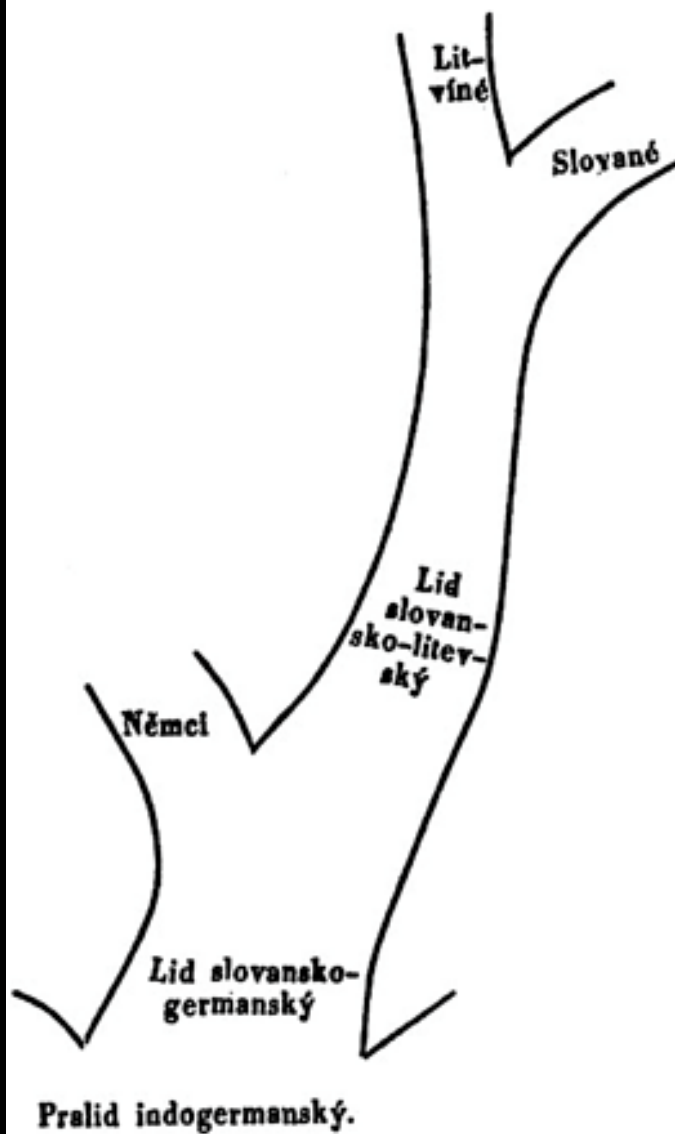
- 言語への系統モデルの適用が流行っている
- でも方言群の語彙に適用するのは無理がある
- 元データを作った人たちは何を考えていた?
- ではまったく違う接触モデルを考えてみる
- シミュレーションデータを系統モデルがどう解釈するか見てみる

本日のおはなし

- 言語への系統モデルの適用が流行っている
- でも方言群の語彙に適用するのは無理がある
- 元データを作った人たちは何を考えていた?
- ではまったく違う接触モデルを考えてみる
- シミュレーションデータを系統モデルがどう解釈するか見てみる

系統モデルの言語への適用

- 19世紀半ば-: 言語学者による言語系統樹
- ~30年前-: 生体情報向け(ベイズ)統計モデル
- ~15年前-: 統計モデルの言語への適用





Partitions Taxa Tips Traits Sites Clocks Trees States Priors Operators MCMC

Taxon Set Mo... St... Tree Age

Tokyo/Kyoto ☐ ☐ ja... ▾

Taxon set: Tokyo/Kyoto

Excluded Taxa

Iwate
Kagawa
Kagoshima
Kanagawa
Kochi
Kumamoto
MJ
Mie
Miyagi
Miyazaki
Nagano
Nagasaki
Nara
Naze
Niigata
OJ
Oita
Okayama
Okinawa
Osaka
Saga
Saitama
Shiga
Shimane
Shizuoka

Select: taxon set... ▾

Included Taxa

Kyoto
Tokyo

Select: taxon set... ▾

+ -

Partitions

Taxa

Tips

Traits

Sites

Clocks

Trees

States

Priors

Operators

MCMC

Taxon Set

Mo...

St...

Tree

Age

Tokyo/Kyoto



ja...



Taxon set: Tokyo/Kyoto

Excluded Taxa

Iwate
Kagawa
Kagoshima
Kanagawa
Kochi
Kumamoto
MJ
Mie
Miyagi
Miyazaki
Nagano
Nagasaki
Nara
Naze
Niigata
OJ
Oita
Okayama
Okinawa
Osaka
Saga
Saitama
Shiga
Shimane
Shizuoka

Select: taxon set...

Included Taxa

Kyoto
Tokyo

Select: taxon set...

プログラミングせ
ずに研究できる

+ -



Partitions Taxa Tips Traits Sites Clocks Trees States Priors Operators MCMC

Taxon Set Mo... St... Tree Age

Tokyo/Kyoto ☐ ☐ ja... ▼

Taxon set: Tokyo/Kyoto

Excluded Taxa

Iwate
Kagawa
Kagoshima
Kanagawa
Kochi
Kumamoto
Mj
Mie
Miyagi
Miyazaki
Nagano
Nagasaki
Nara
Naze
Niigata
Oj
Oita
Okayama
Okinawa
Osaka
Saga
Saitama
Shiga
Shimane
Shizuoka

Select: taxon set... ▼

Included Taxa

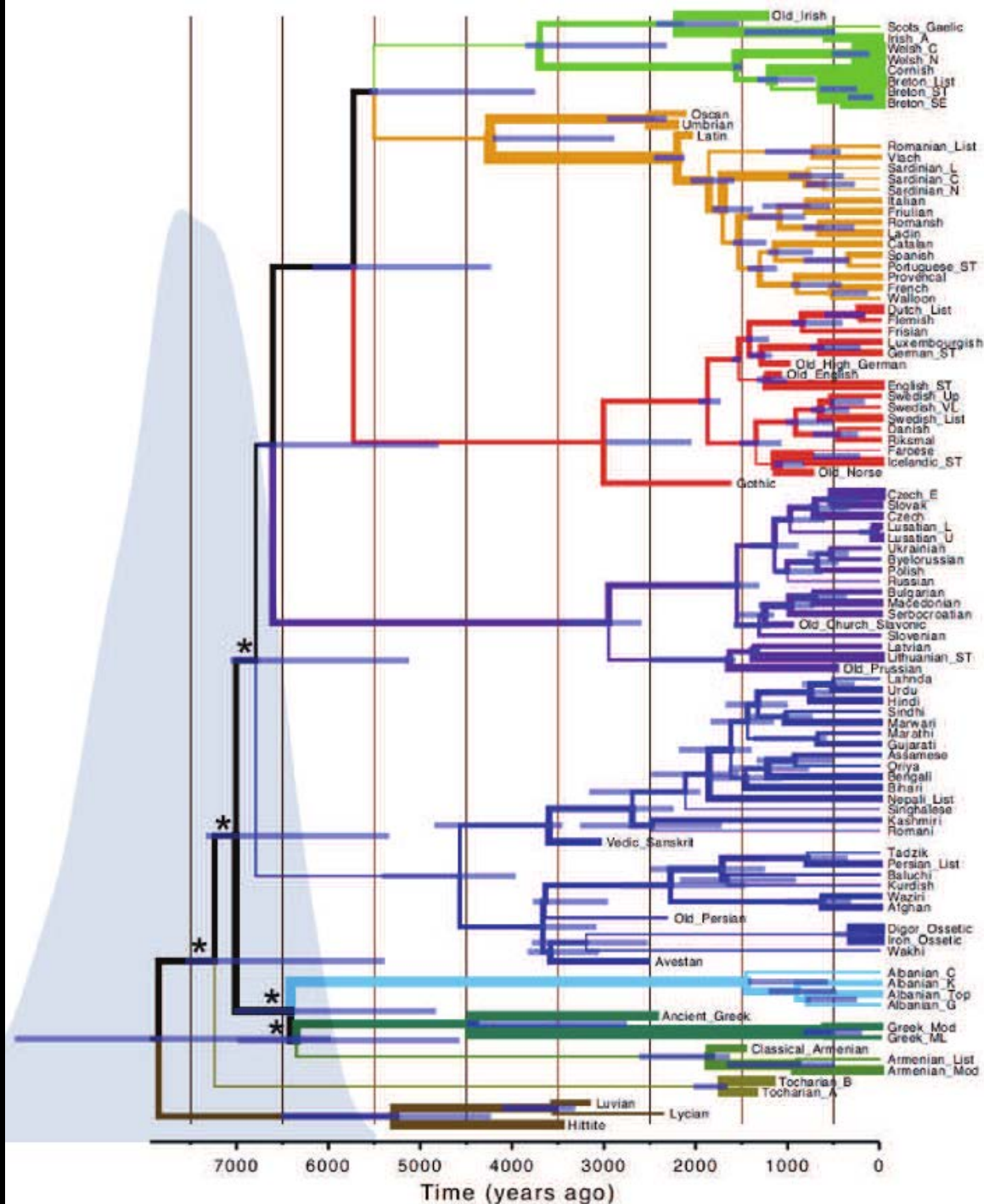
Kyoto
kyo

Select: taxon set... ▼

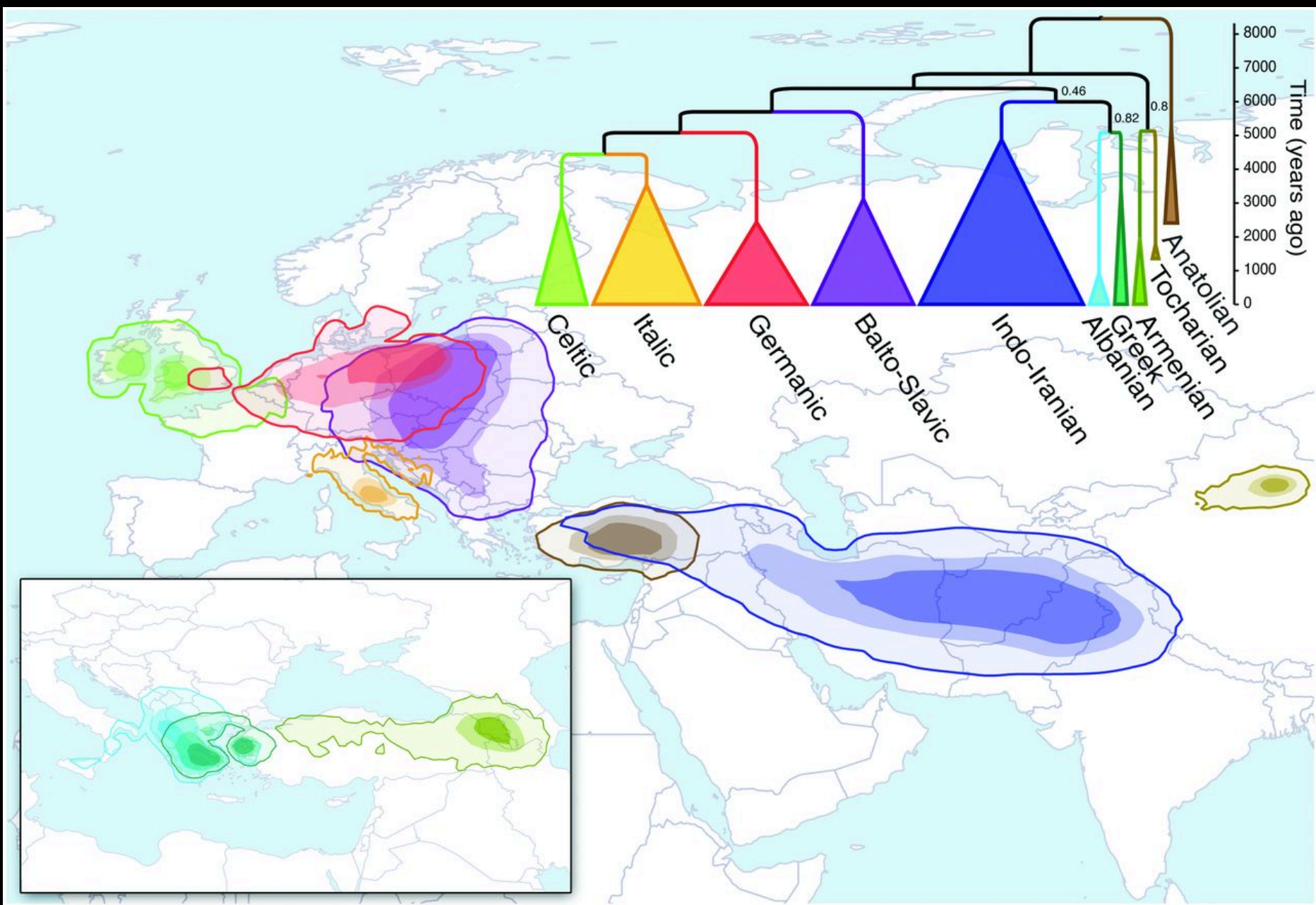
+ -

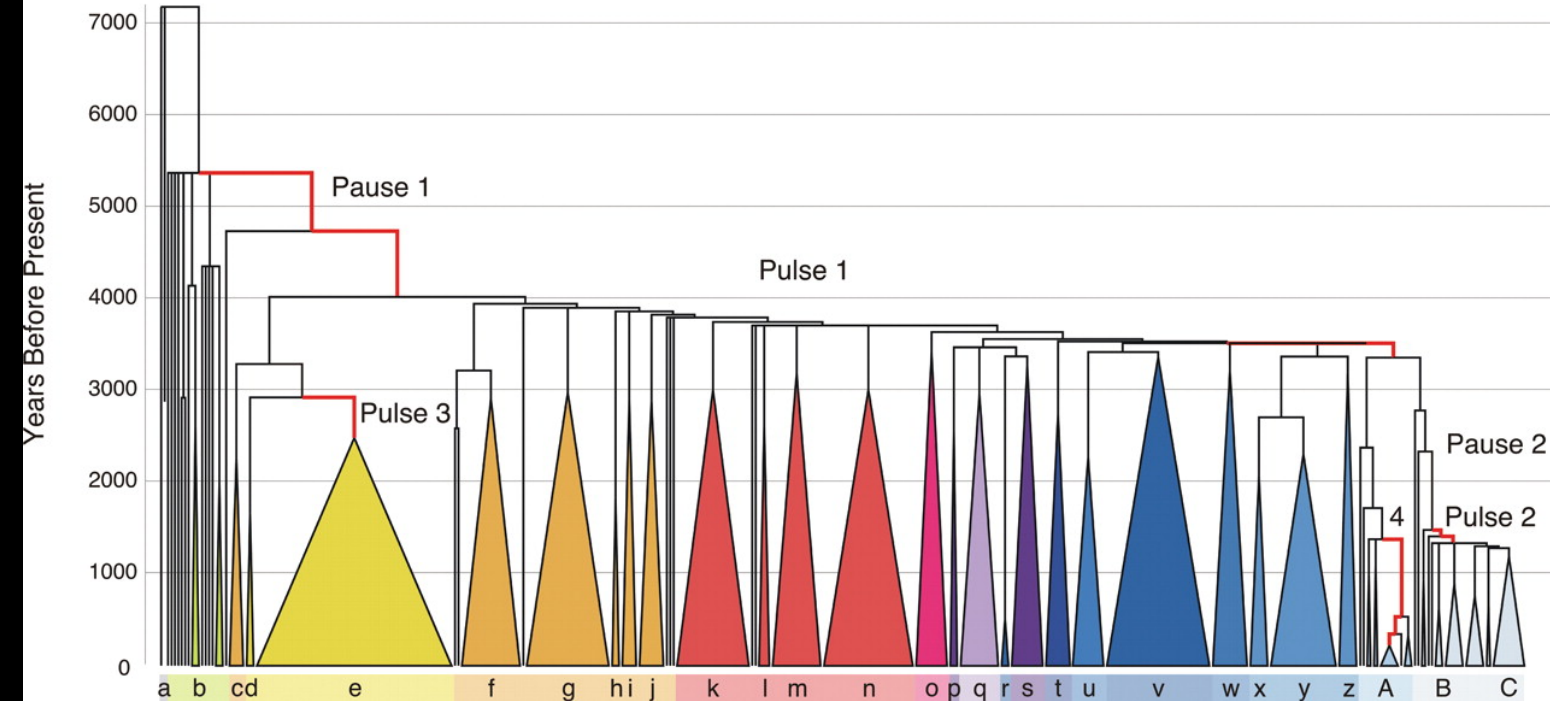
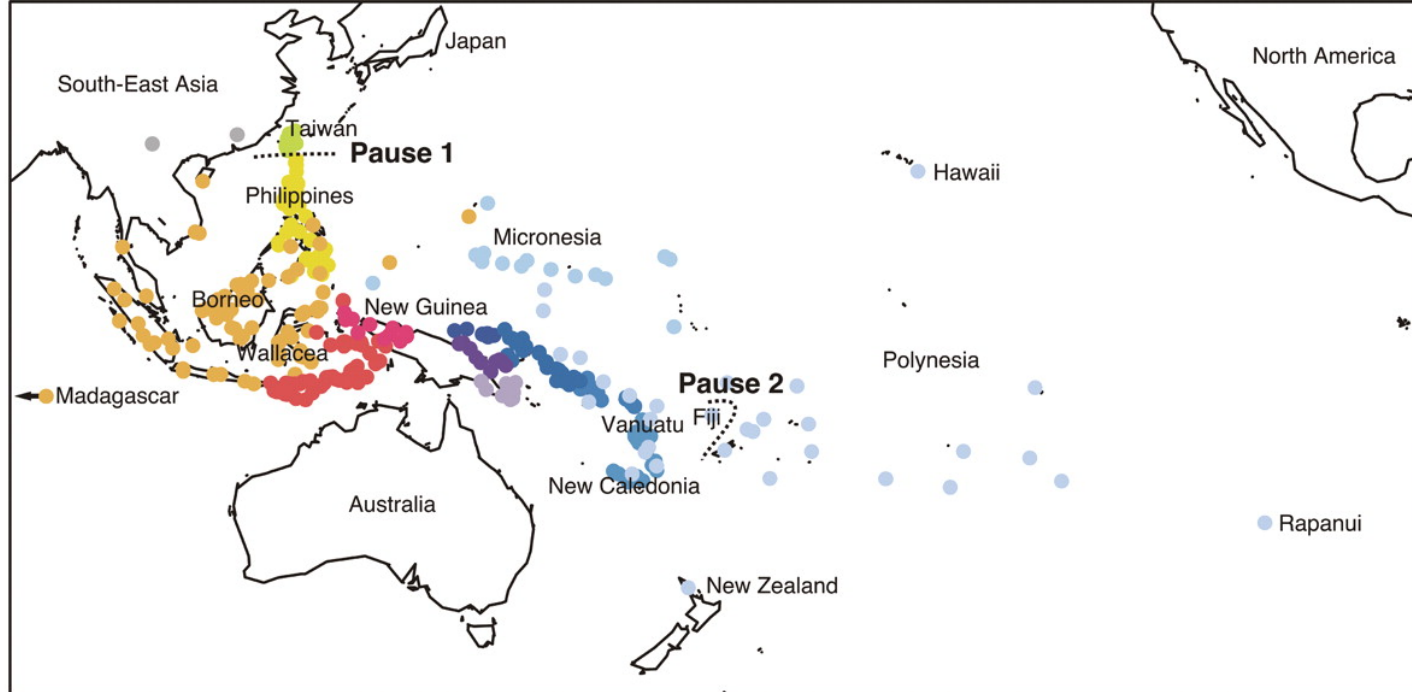
プログラミングせ
ずに研究できる

言語屋 = データ屋?

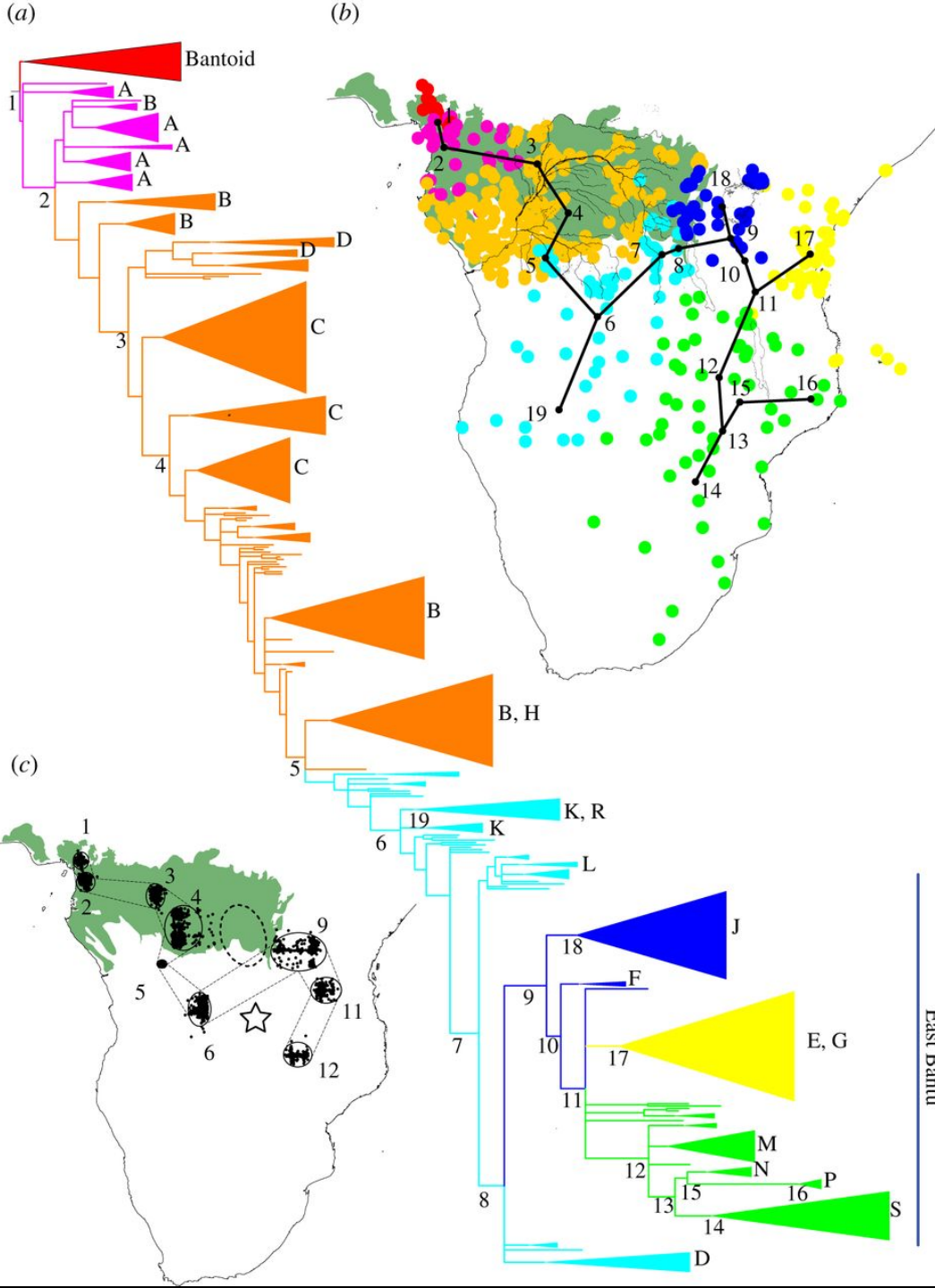


[Bouckaert+, 2012]
(Science)





[Gray+, 2012]
(Science)



[Currie+, 2013]
(R. Soc. B)

ノード (各言語) は基礎語彙

水

大きい

英語

water

big

ドイツ語

Wasser

gross

ロシア語

вода

большой великий

フランス語

eau

grand

イタリア語

acqua

grande

ノード (各言語) は基礎語彙

水

大きい

英語

water

big

ドイツ語

Wasser

gross

ロシア語

вода

большой

великий

フランス語

eau

grand

イタリア語

acqua

grande

ノード (各言語) は基礎語彙

水

大きい

英語

1

water

3

big

ドイツ語

Wasser

4

gross

ロシア語

вода

5

большой

6

великий

フランス語

2

eau

7

grand

イタリア語

acqua

grande

ノード (各言語) は基礎語彙

	水	大きい	
英語	¹ water	³ big	{1, 3}
ドイツ語	Wasser	⁴ gross	{1, 4}
ロシア語	вода	⁵ большой	⁶ великий {1, 5, 6}
フランス語	² eau	⁷ grand	{2, 7}
イタリア語	acqua	grande	{2, 7}

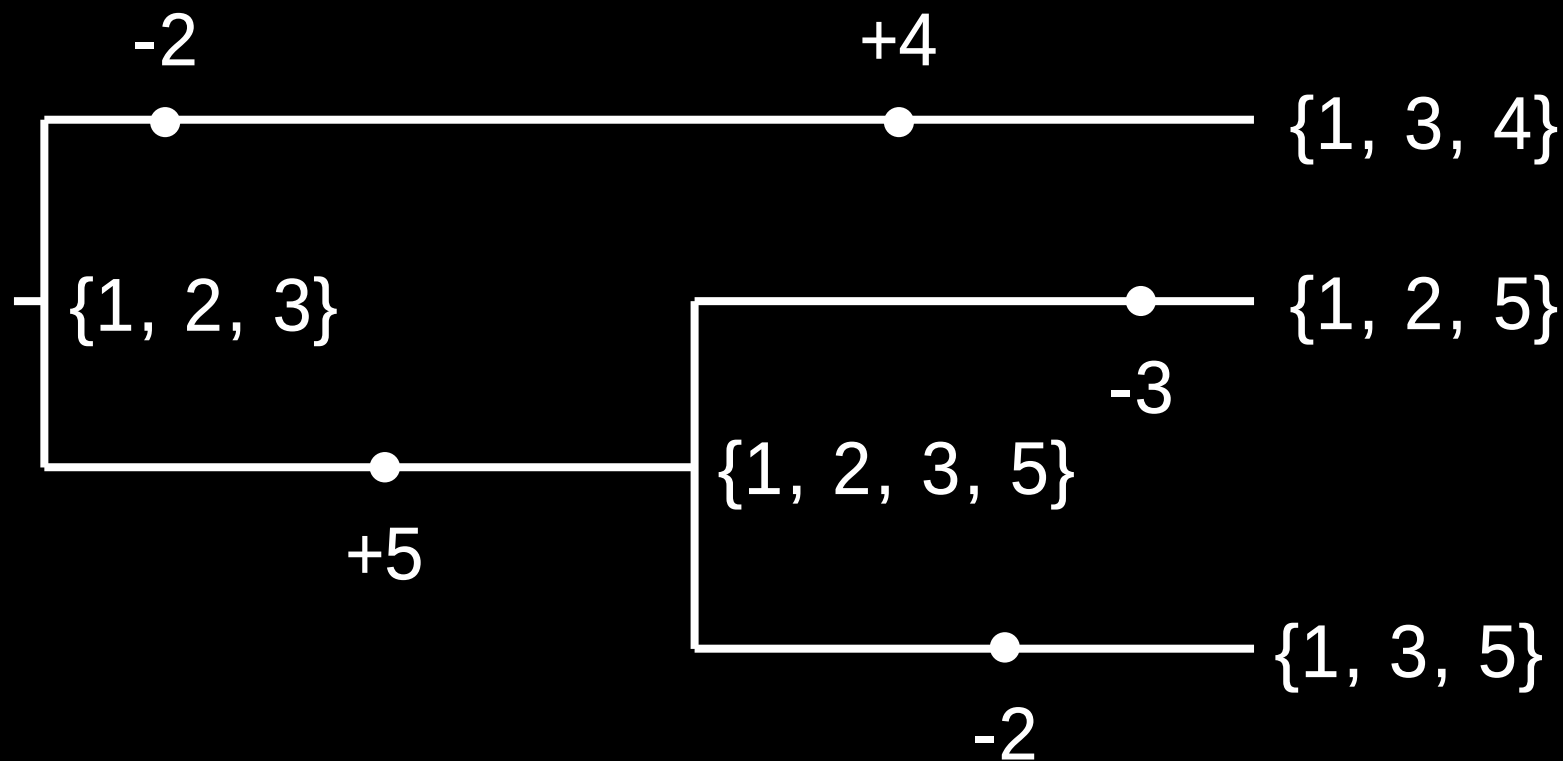
木を特徴づけるのは分岐

$\{1, 3, 4\}$

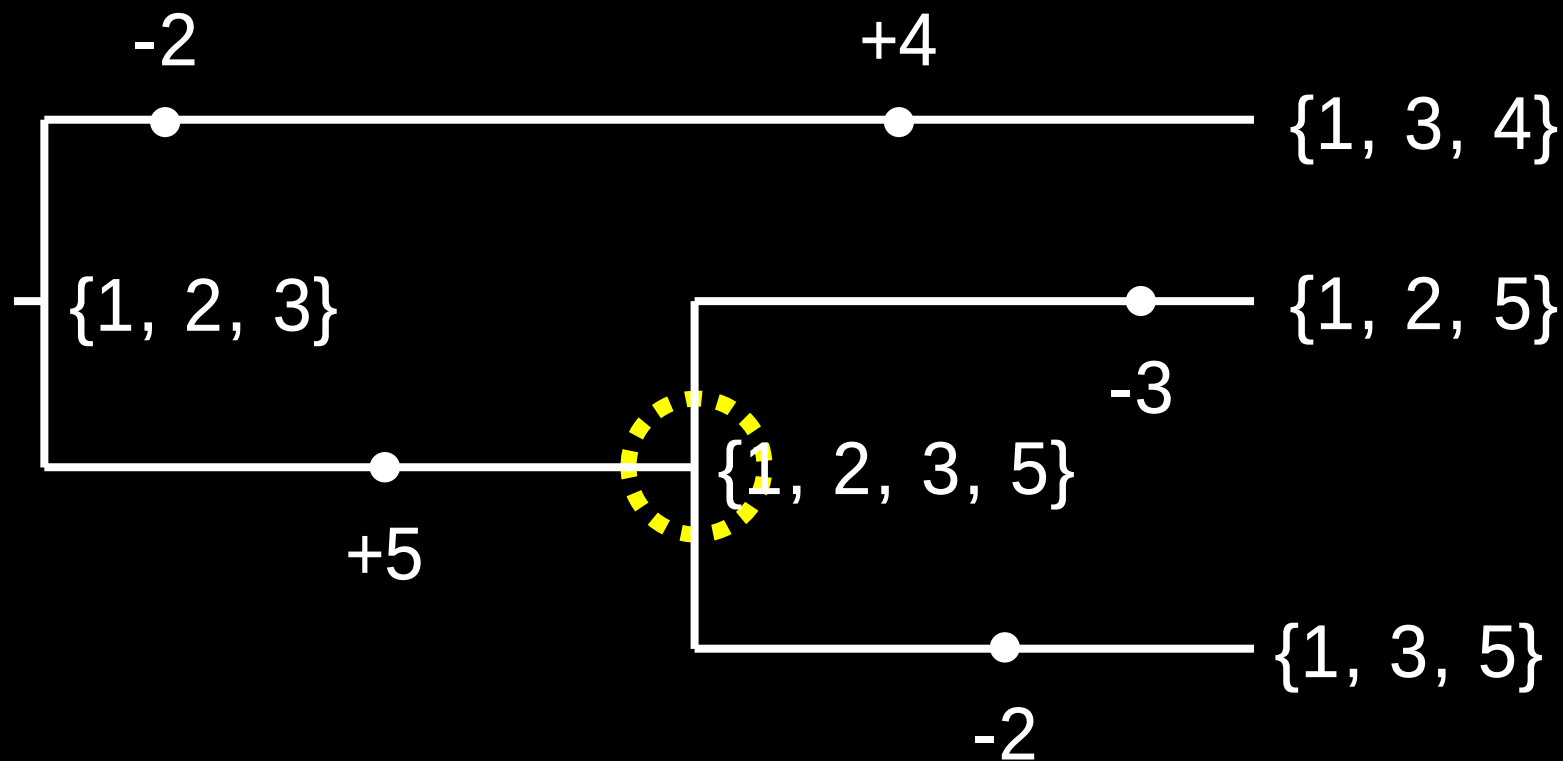
$\{1, 2, 5\}$

$\{1, 3, 5\}$

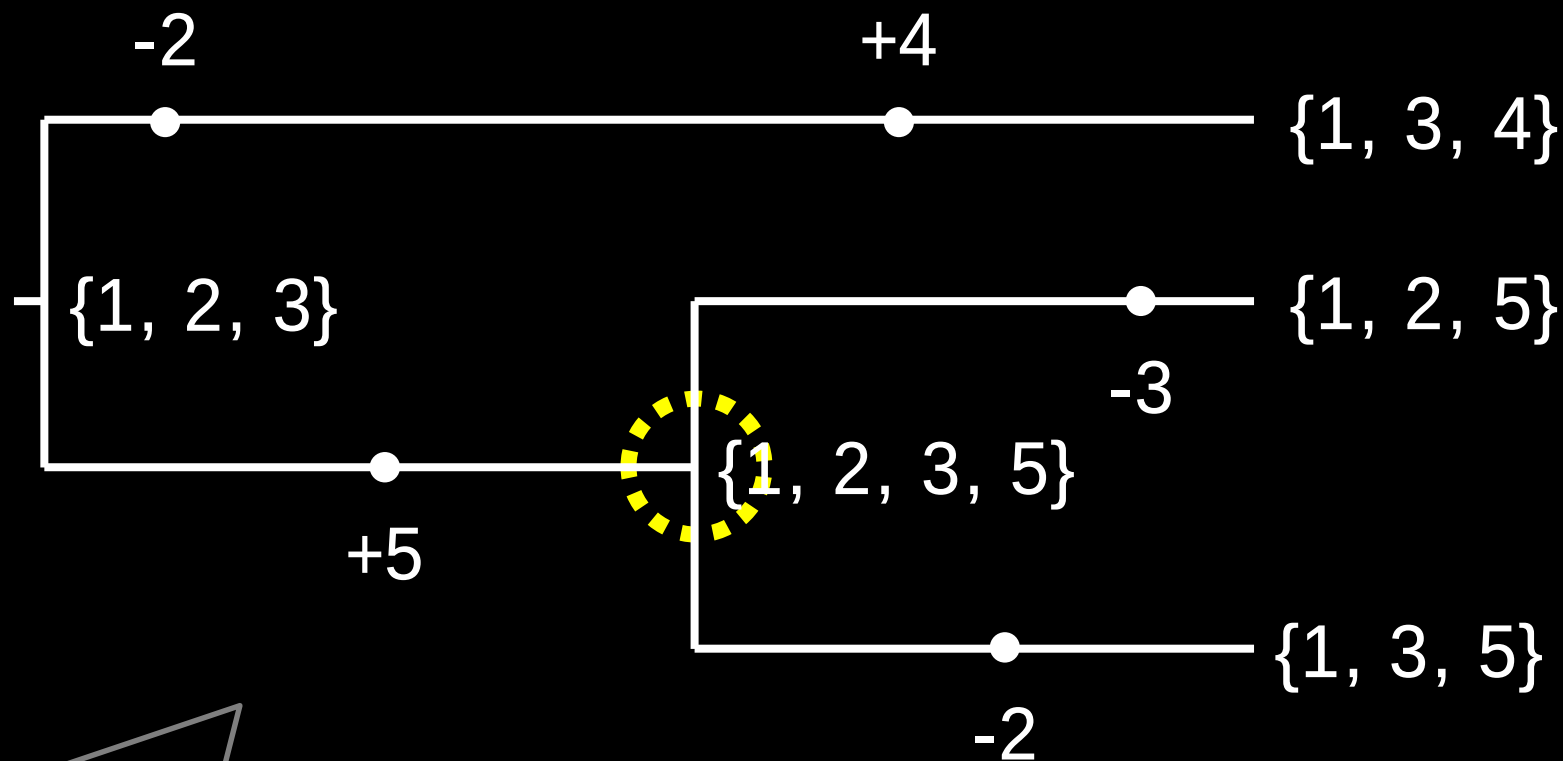
木を特徴づけるのは分岐



木を特徴づけるのは分岐

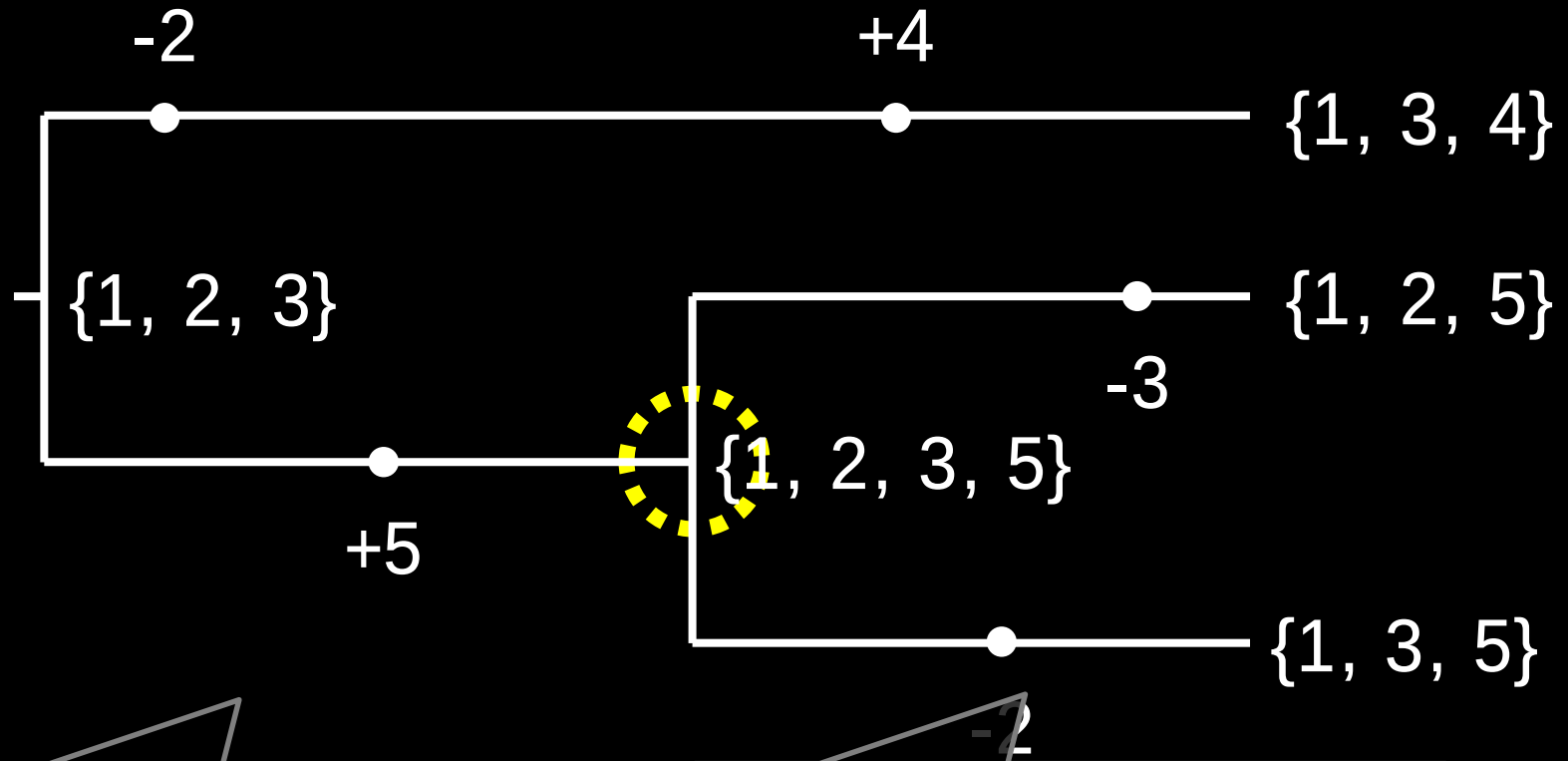


木を特徴づけるのは分岐



分岐前は
同一の状態

木を特徴づけるのは分岐

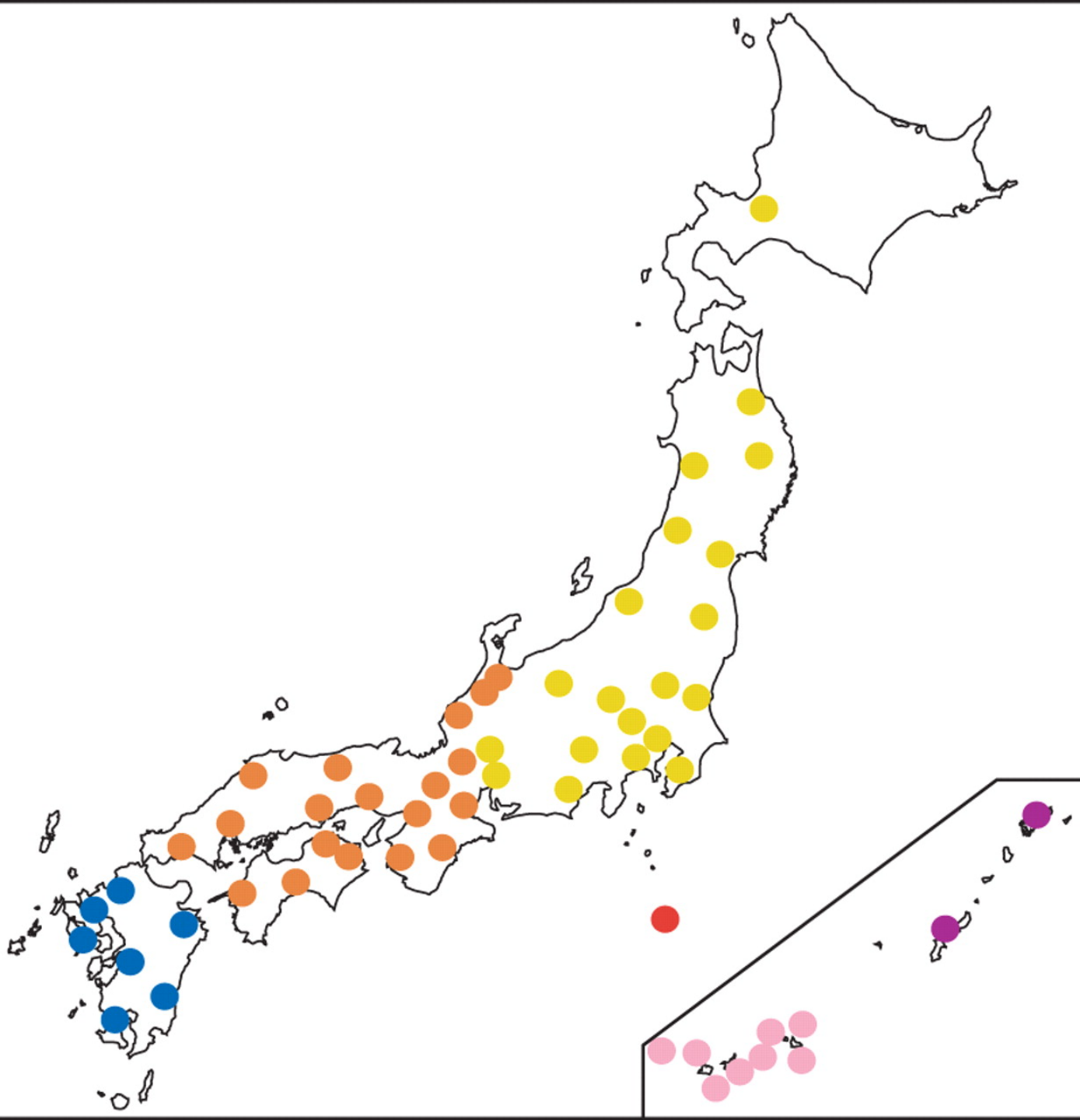


分岐前は
同一の状態

分岐後は
別々に変化

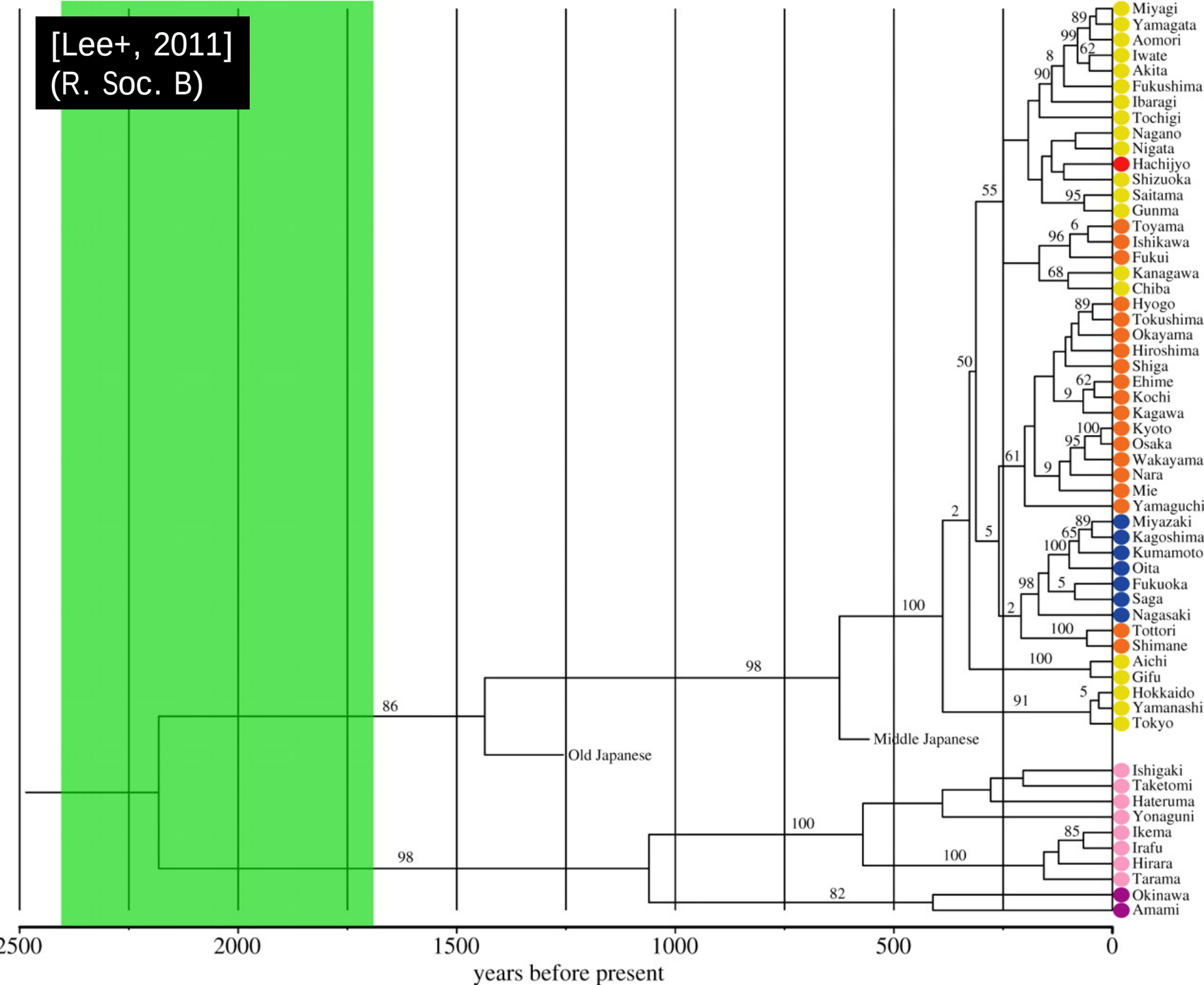
本日のおはなし

- 言語への系統モデルの適用が流行っている
- でも方言群の語彙に適用するのは無理がある
- 元データを作った人たちは何を考えていた?
- ではまったく違う接触モデルを考えてみる
- シミュレーションデータを系統モデルがどう解釈するか見てみる

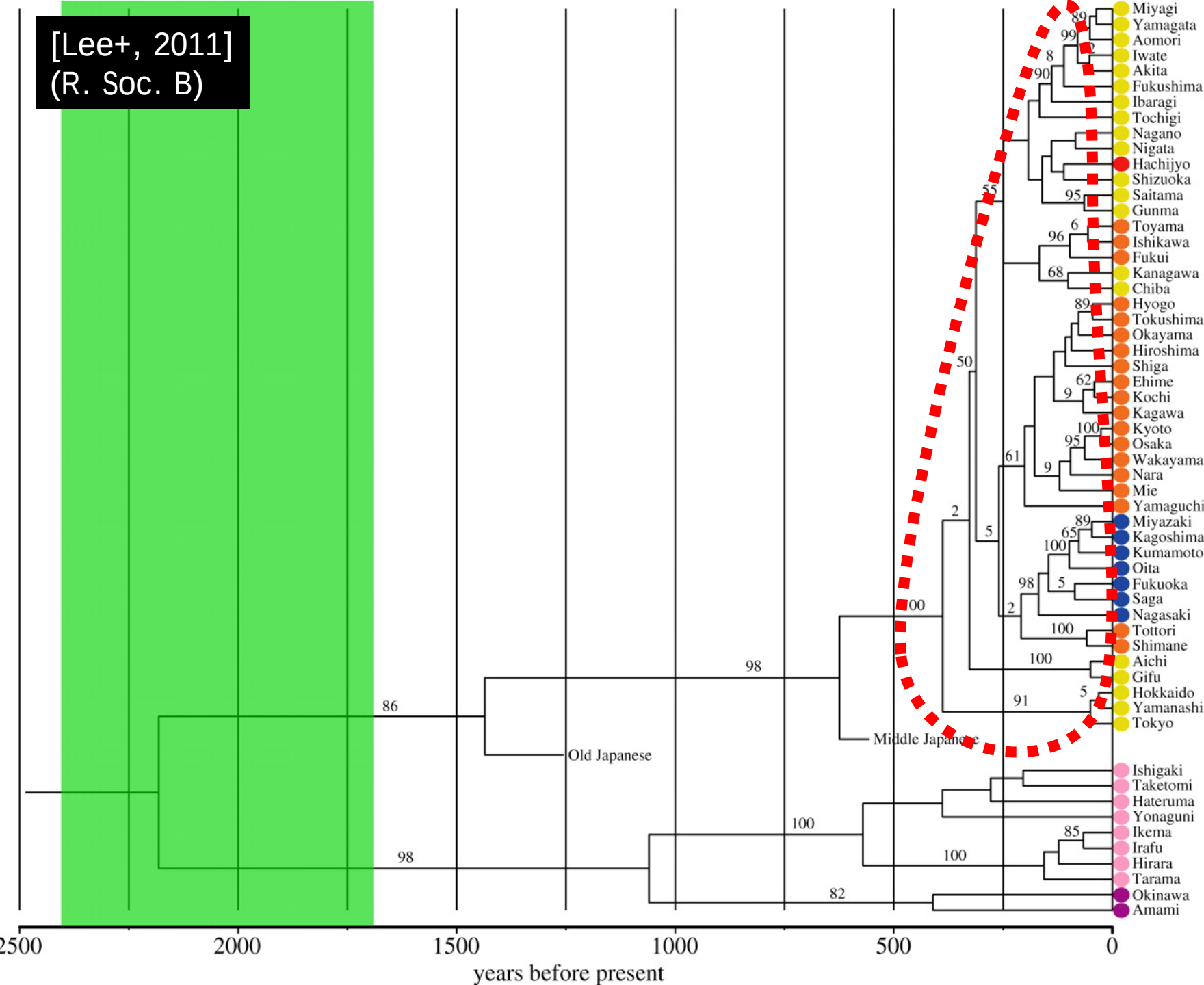


[Lee+, 2011]
(R. Soc. B)

[Lee+, 2011]
(R. Soc. B)

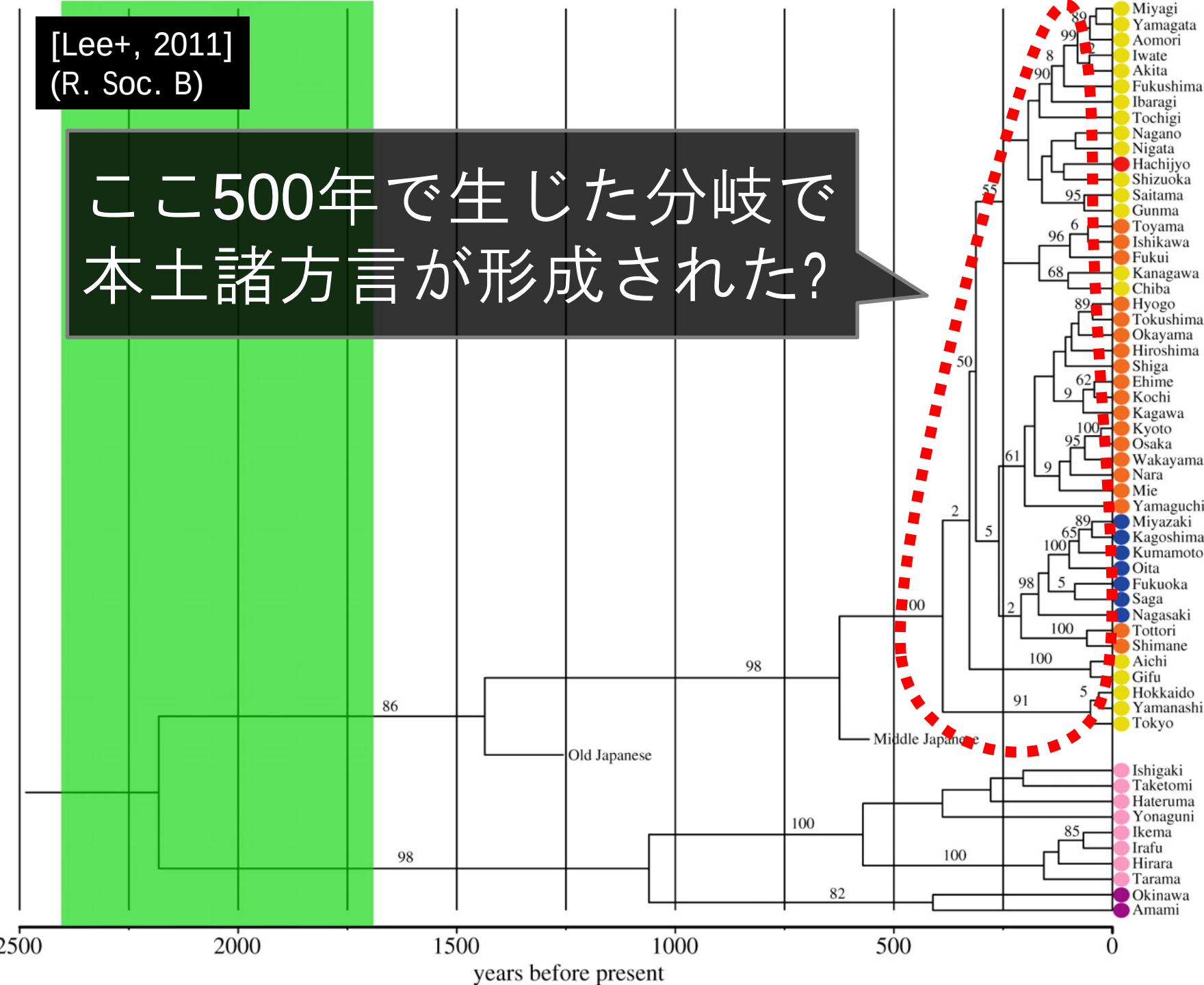


[Lee+, 2011]
(R. Soc. B)



[Lee+, 2011]
(R. Soc. B)

ここ500年で生じた分岐で
本土諸方言が形成された?

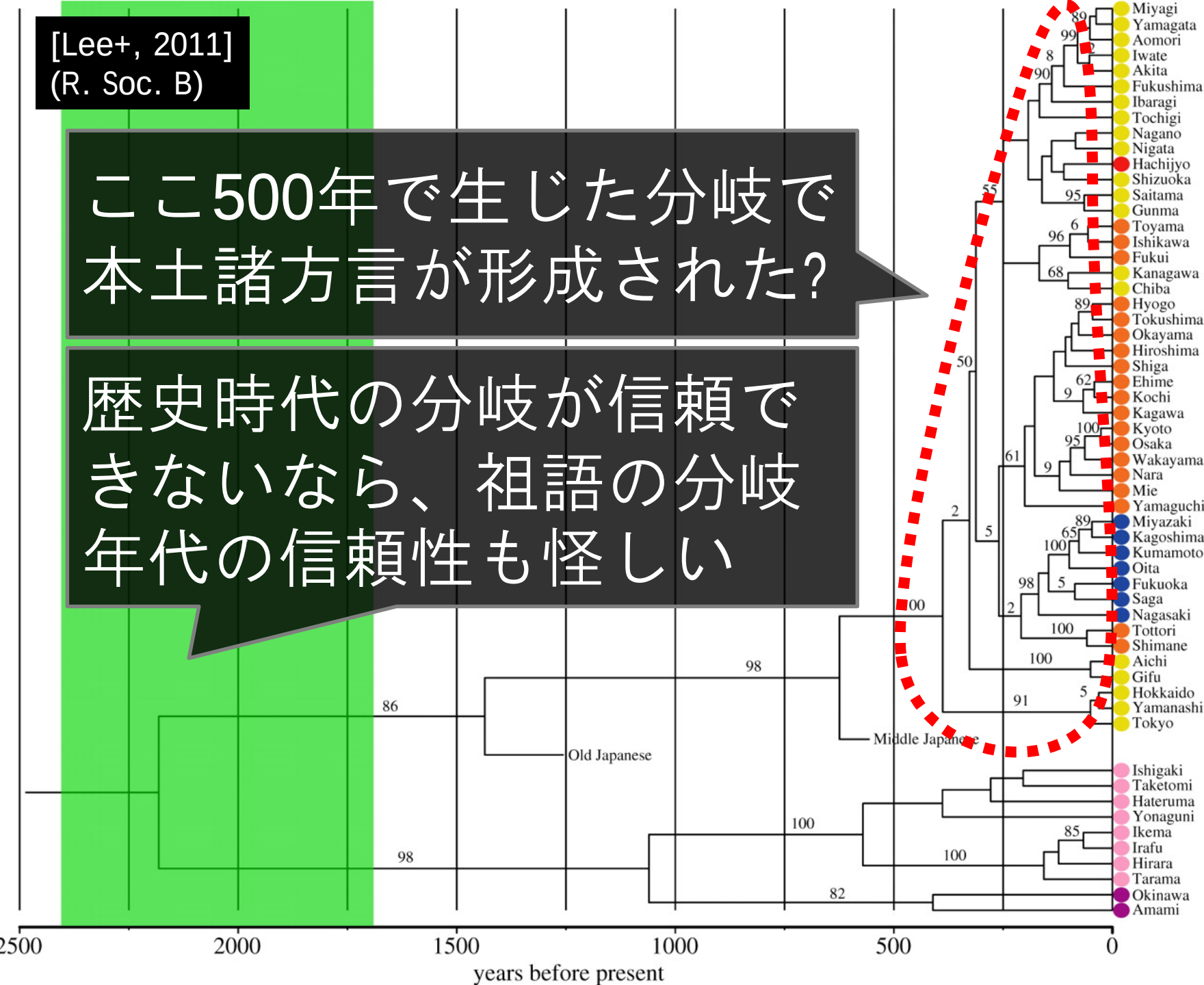


[Lee+, 2011]

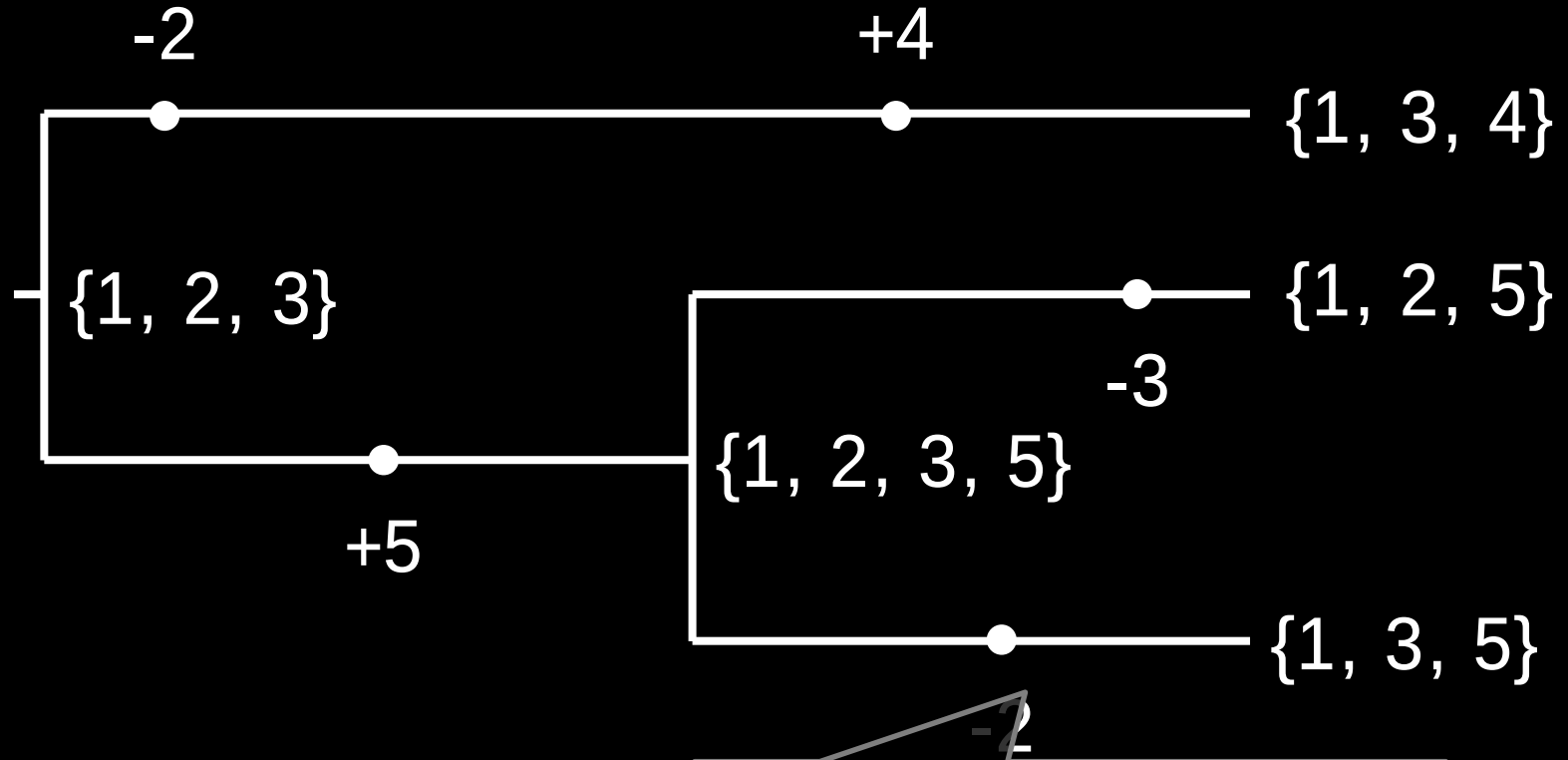
(R. Soc. B)

ここ500年で生じた分岐で
本土諸方言が形成された？

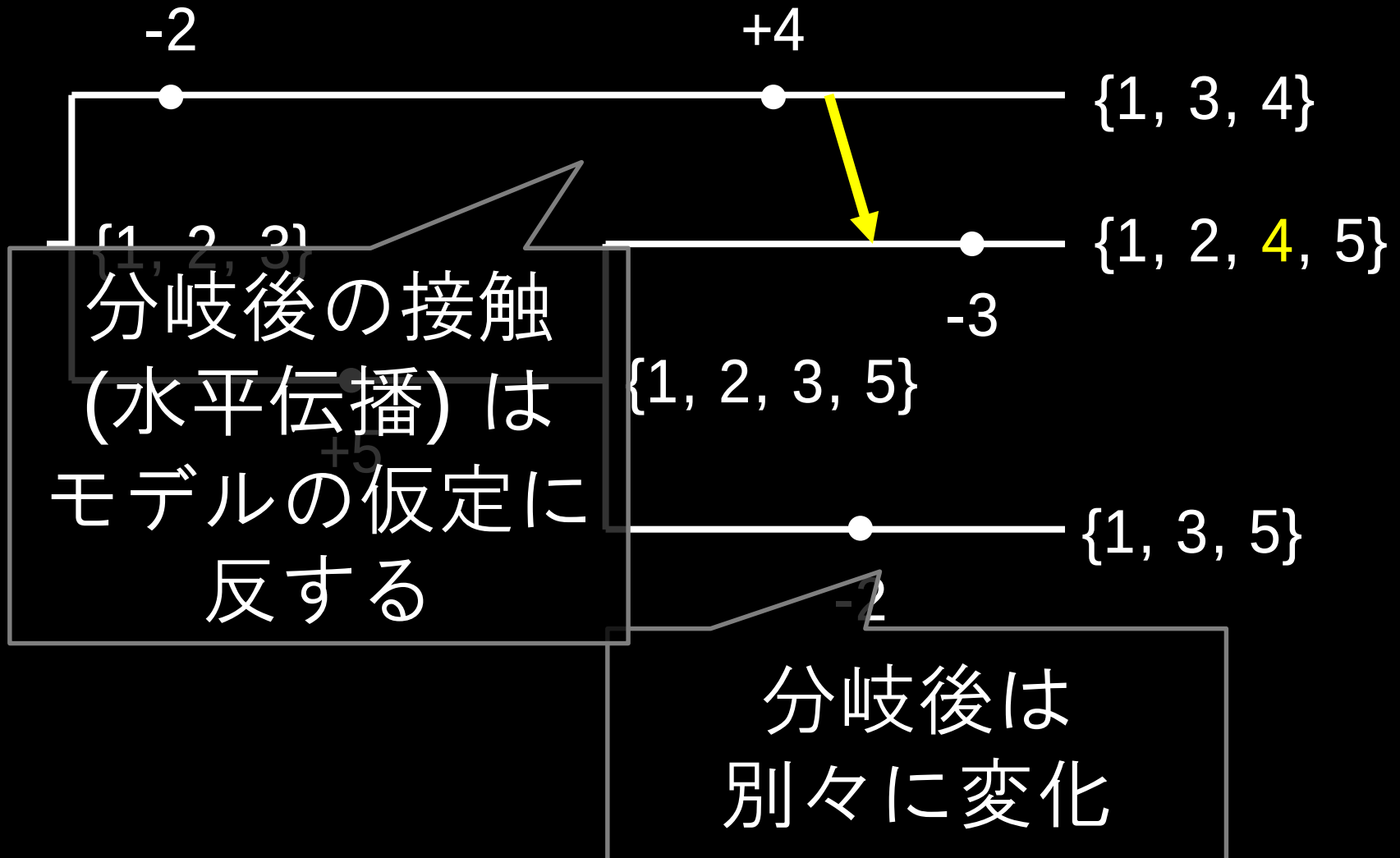
歴史時代の分岐が信頼で
きないなら、祖語の分岐
年代の信頼性も怪しい

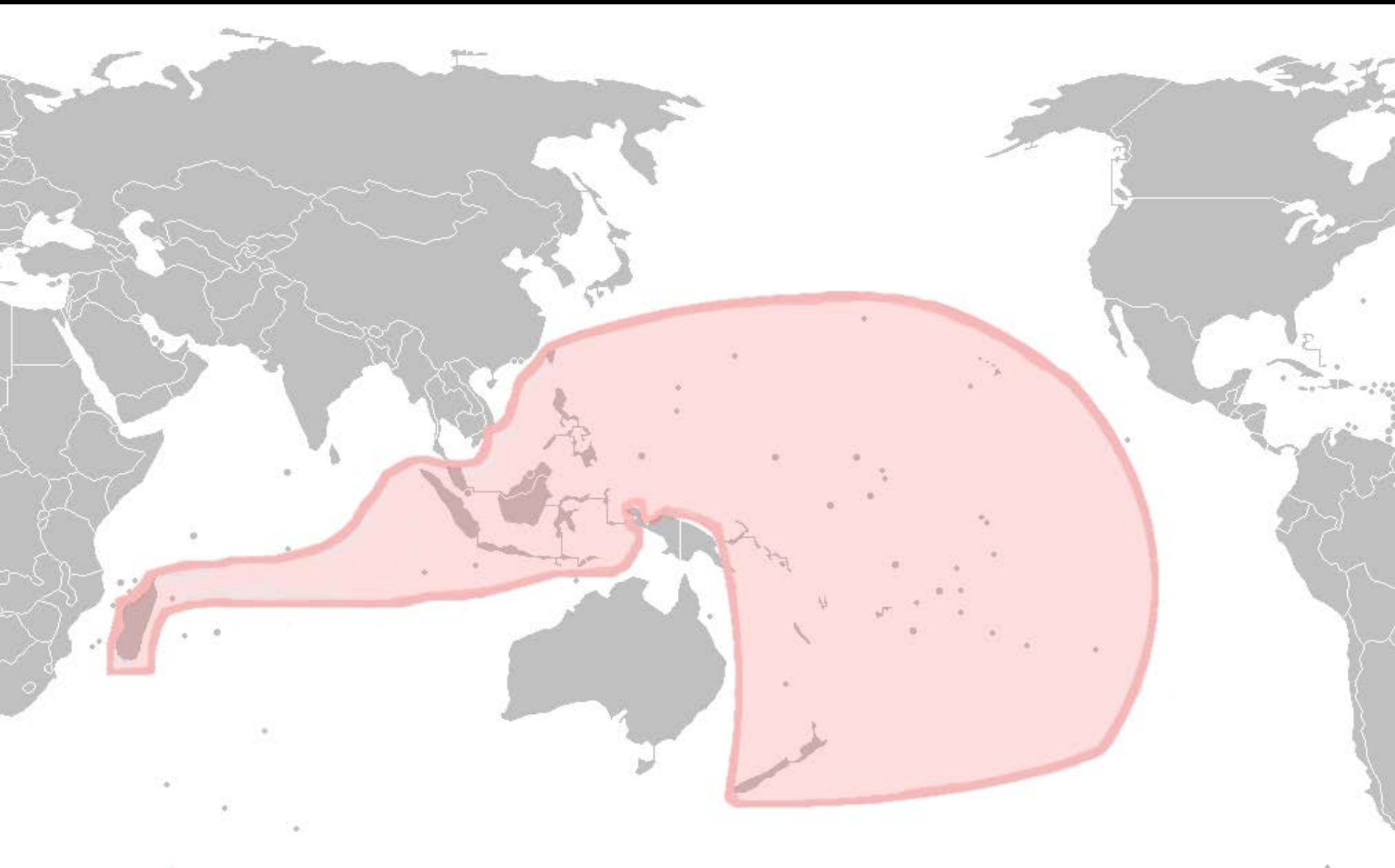


木を特徴づけるのは分岐 (再掲)

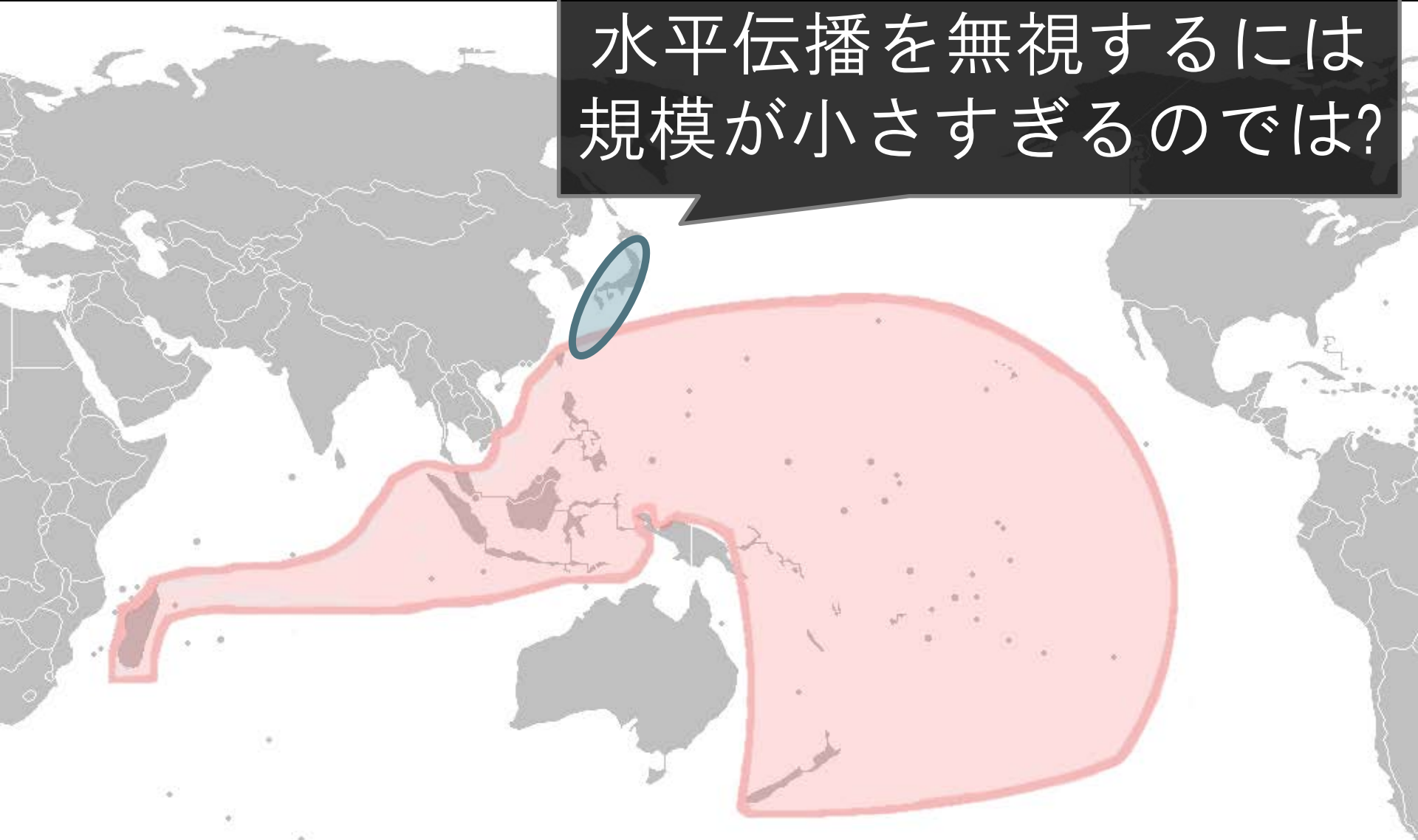


木を特徴づけるのは分岐 (再掲)





水平伝播を無視するには
規模が小さすぎるのでは？



本日のおはなし

- 言語への系統モデルの適用が流行っている
- でも方言群の語彙に適用するのは無理がある
- 元データを作った人たちは何を考えていた?
- ではまったく違う接触モデルを考えてみる
- シミュレーションデータを系統モデルがどう解釈するか見てみる

言語年代学 (Glottochronology)

- スワデシュ (1909-1967) が発展させた統計手法
 - 仮定1: 安定的な語彙 (基礎語彙) が存在
 - 仮定2: 基礎語彙の置き換えは一定割合で進む
 - 言語ペアの基礎語彙 + 置き換え率 $\Rightarrow$ 分岐年代

言語年代学 (Glottochronology)

- スワデシュ (1909-1967) が発展させた統計手法
 - 仮定1: 安定的な語彙 (基礎語彙) が存在
 - 仮定2: 基礎語彙の置き換えは一定割合で進む
 - 言語ペアの基礎語彙 + 置き換え率 $\Rightarrow$ 分岐年代
- 基礎語彙データベース
 - [IELex](#): インド・ヨーロッパ語族
 - [Austronesian Basic Vocabulary Database](#)
 - アイヌ語諸方言基礎語彙 [服部+, 1960]

言語年代学 (Glottochronology)

- スワデシュ (1909-1967) が発展させた統計手法
 - 仮定1: 安定的な語彙 (基礎語彙) が存在
 - 仮定2: 基礎語彙の置き換えは一定割合で進む
 - 言語ペアの基礎語彙 + 置き換え率 $\Rightarrow$ 分岐年代
- 基礎語彙データベース
 - [IELex](#): インド・ヨーロッパ語族
 - [Austronesian Basic Vocabulary Database](#)
 - アイヌ語諸方言基礎語彙 [服部+, 1960]
 - 日本語?

基礎語彙調査、方言地理学

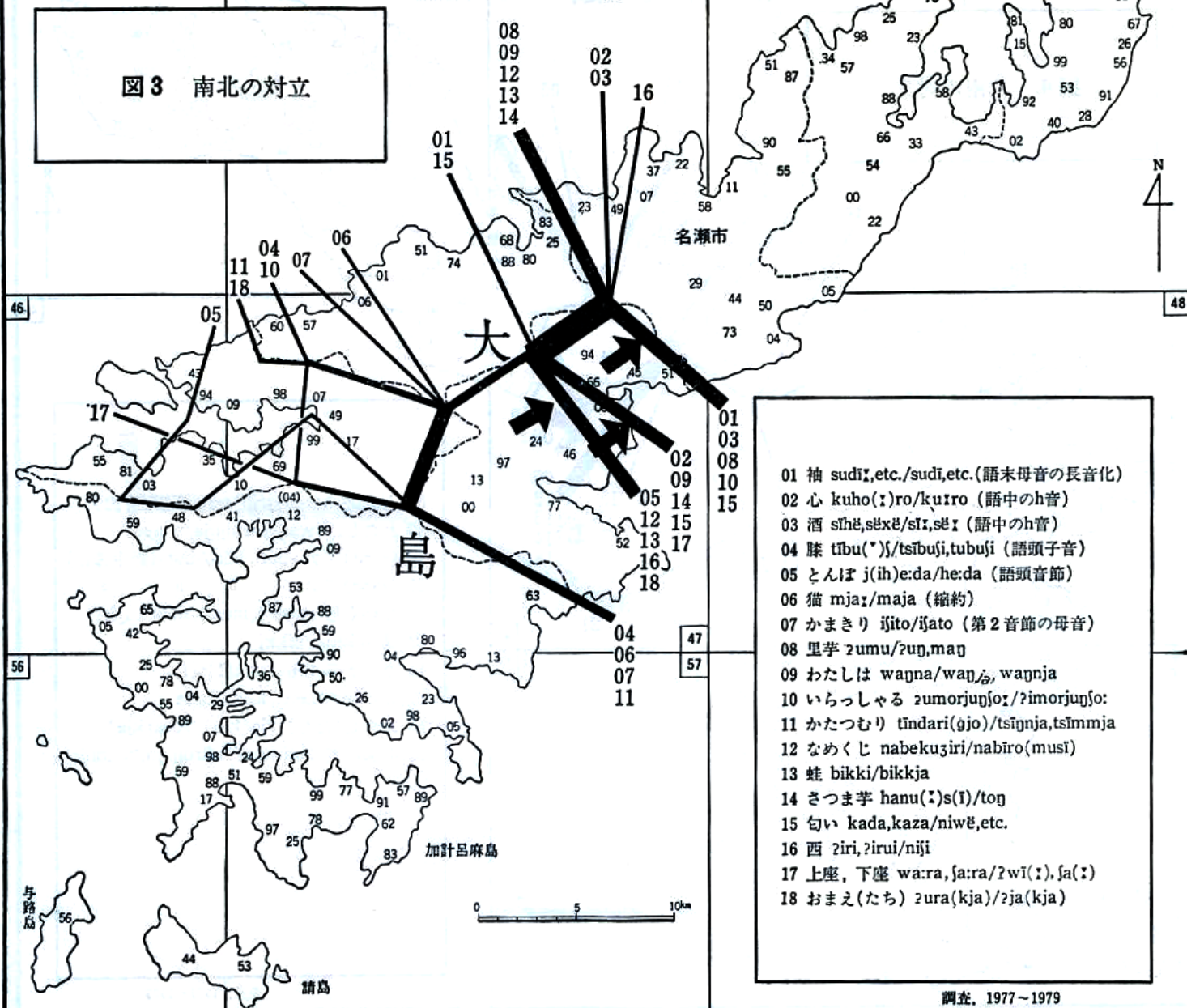
- 日本語データの主な出典:
『現代日本方言大辞典』
(1992-1994)
- 言語年代学とは別系統の研究
- 100年以上の研究の蓄積
 - Atlas linguistique de la France (1902-1912)
 - 『音韻調査報告書』、『音韻分布図』 (1905)



奄美大島言語地図

柴田武・日下部文夫・真田信治
沢木幹栄・下野雅昭・三石泰子

図3 南北の対立

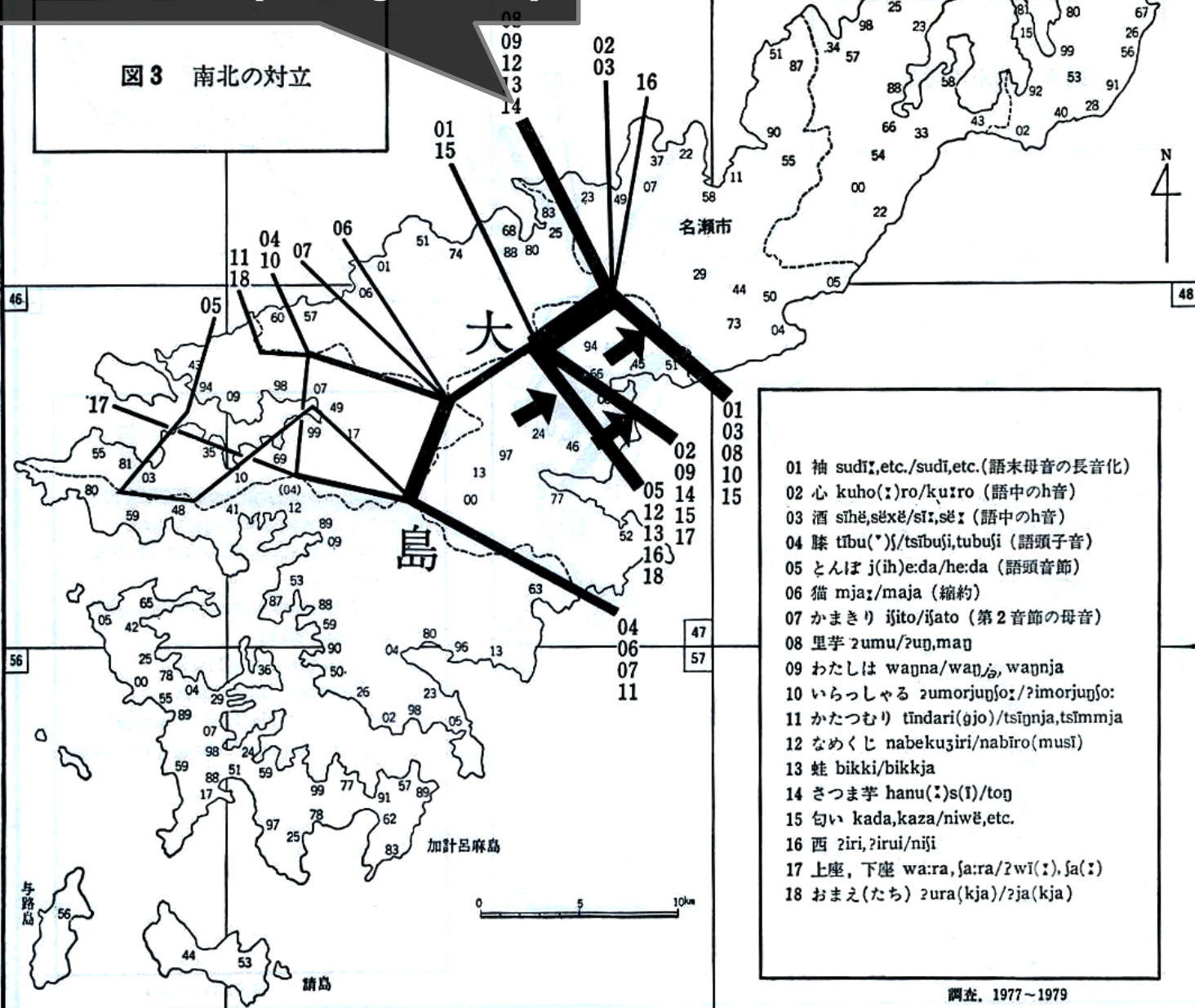


[柴田, 1982]

奄美大島言語地図

等語線 (isogloss)

図3 南北の対立



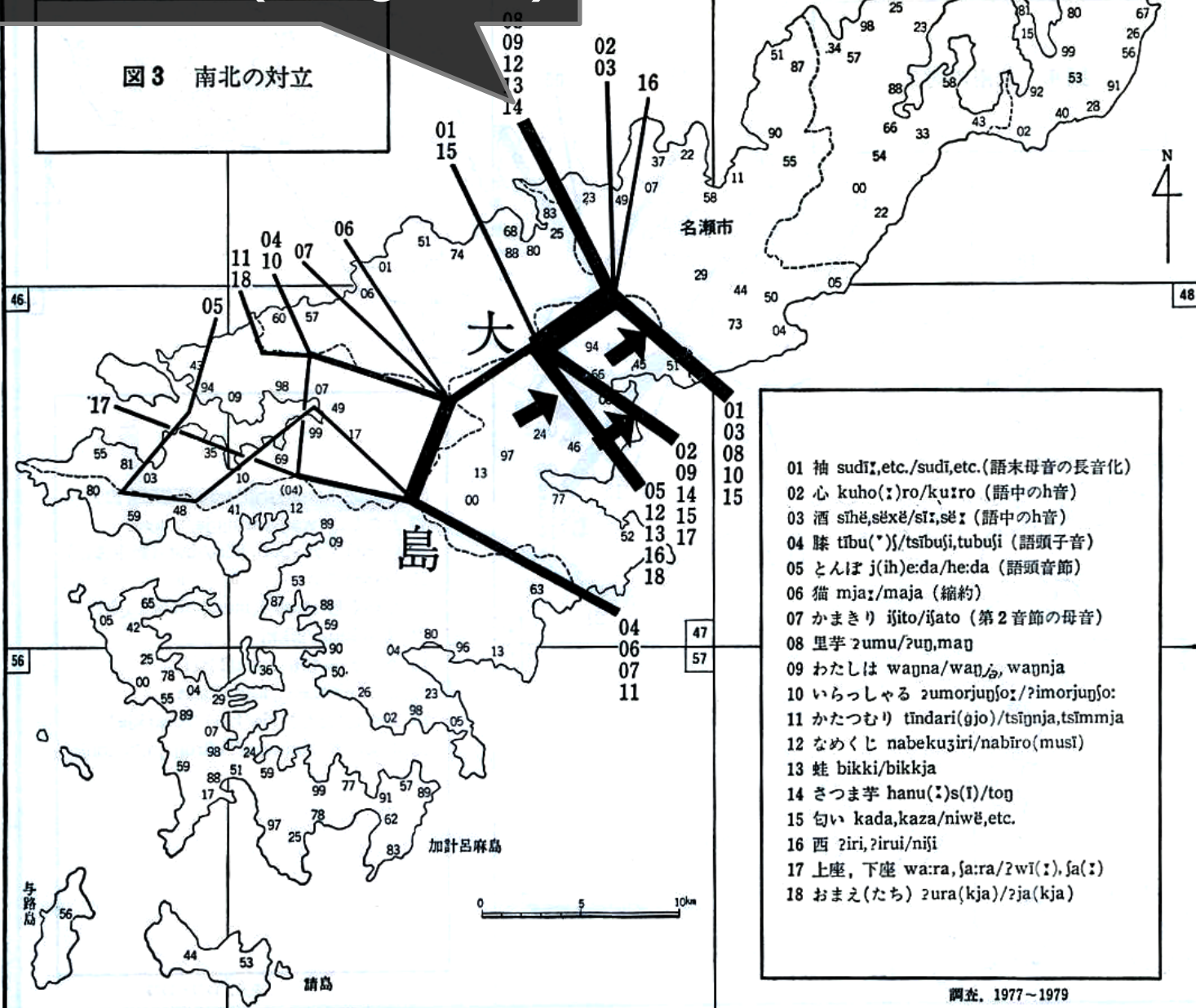
[柴田, 1982]

奄美大島言語地図

等語線 (isogloss)

スナップシヨット

図3 南北の対立



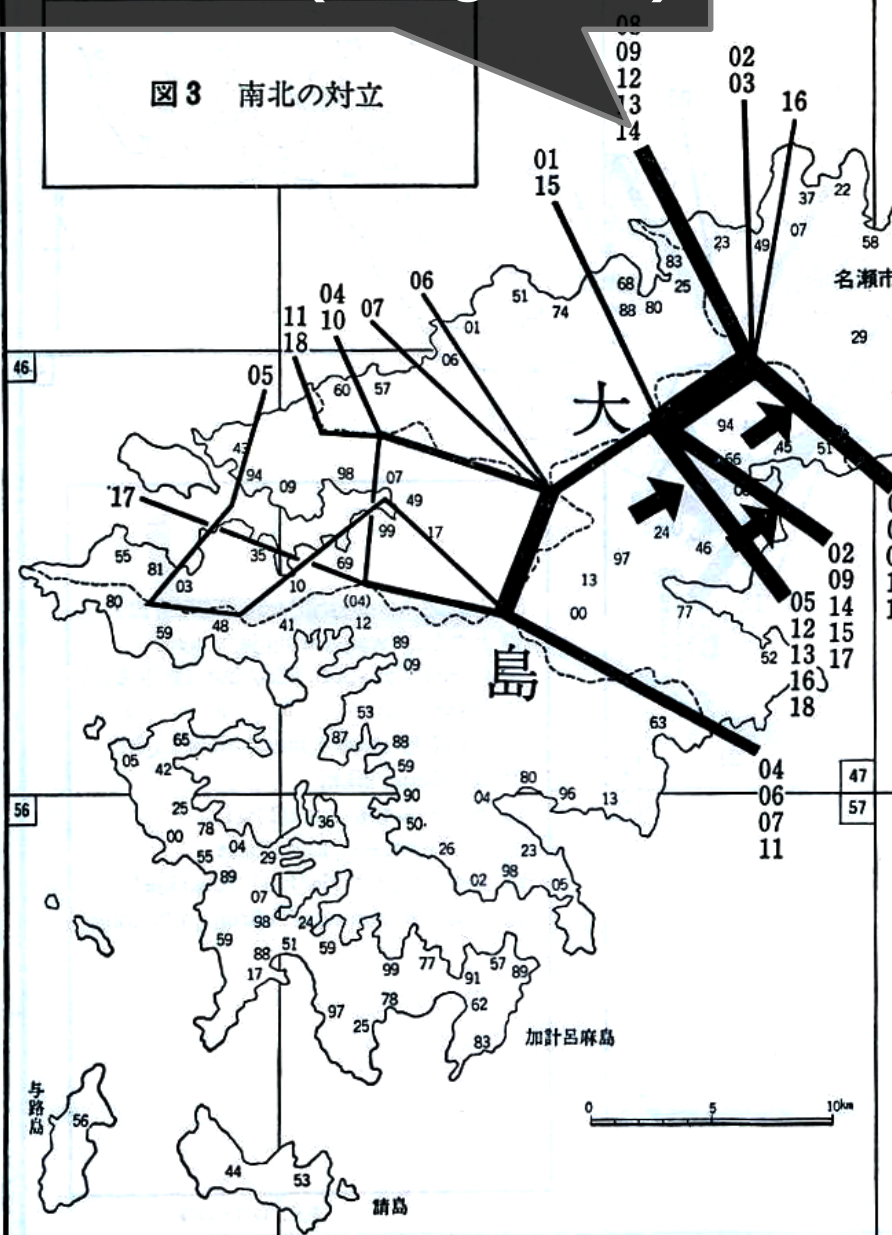
- 01 袖 *sudĩ*, etc./*sudĩ*, etc. (語末母音の長音化)
- 02 心 *kuho(ː)ro/kuːro* (語中のh音)
- 03 酒 *sĩhẽ,sẽxẽ/sĩː,sẽː* (語中のh音)
- 04 膝 *tĩbu(ː)/tsĩbuːi,tubuːi* (語頭子音)
- 05 とんぼ *j(ih)e:da/he:da* (語頭音節)
- 06 猫 *mjaː/maja* (縮約)
- 07 かまきり *ĩjito/ĩjato* (第2音節の母音)
- 08 里芋 *ʔumu/ʔug,mag*
- 09 わたしは *wagna/wap,ə, wagnja*
- 10 いらっしゃる *ʔumorjũŋoː/ʔimorjũŋoː*
- 11 かたつむり *tĩdari(gjo)/tsĩgnja,tsĩmmja*
- 12 なめくじ *nabekũjiri/nabiːro(musĩ)*
- 13 蛙 *bikki/bikkja*
- 14 さつま芋 *hanu(ː)s(I)/toŋ*
- 15 匂い *kada,kaza/niwẽ,etc.*
- 16 西 *ʔiri,ʔirui/nĩi*
- 17 上座, 下座 *wa:ra,ʃa:ra/ʔwĩ(ː),ʃa(ː)*
- 18 おまえ(たち) *ʔura(kja)/ʔja(kja)*

[柴田, 1982]

奄美大島言語地図

等語線 (isogloss)

図3 南北の対立



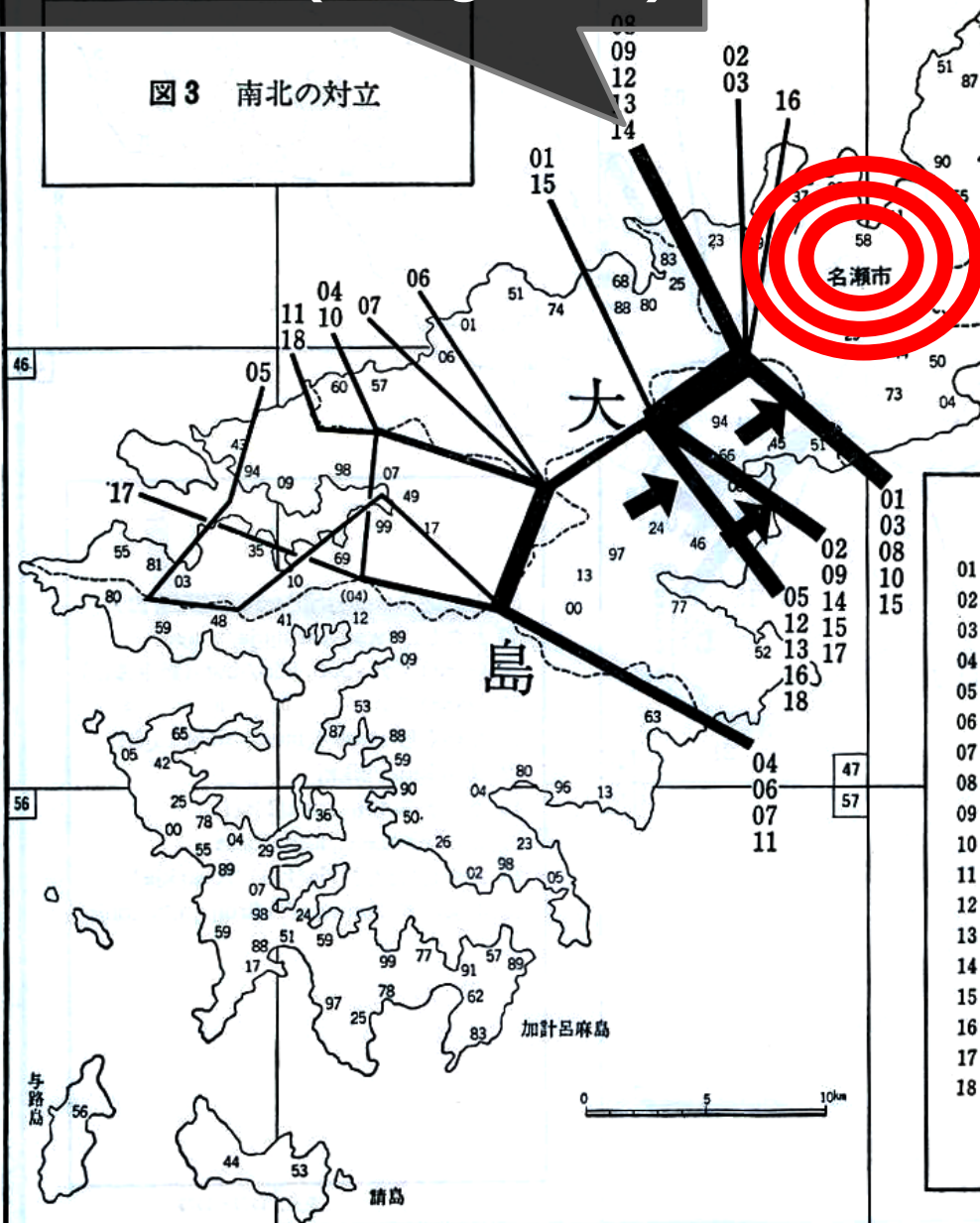
スナップショット 系統上の分岐?

- 01 袖 *sudĩ*, etc./*sudĩ*, etc. (語末母音の長音化)
- 02 心 *kuho(ː)ro/kuːro* (語中のh音)
- 03 酒 *sĩhẽ,sẽxẽ/sĩː,sẽː* (語中のh音)
- 04 膝 *tĩbu(ː)/tsĩbuːi,tubuːi* (語頭子音)
- 05 とんぼ *j(ih)e:da/he:da* (語頭音節)
- 06 猫 *mjaː/maja* (縮約)
- 07 かまきり *ĩjito/ĩjato* (第2音節の母音)
- 08 里芋 *ʔumu/ʔug,mag*
- 09 わたしは *wagna/wap,ə, wagnja*
- 10 いらっしゃる *ʔumorjũŋoː/ʔimorjũŋoː*
- 11 かたつむり *tĩdari(gjo)/tsĩgnja,tsĩmmja*
- 12 なめくじ *nabekũjiri/nabĩro(musĩ)*
- 13 蛙 *bikki/bikkja*
- 14 さつま芋 *hanu(ː)s(I)/toŋ*
- 15 匂い *kada,kaza/niwẽ,etc.*
- 16 西 *ʔiri,ʔirui/nĩji*
- 17 上座, 下座 *wa:ra,ʃa:ra/ʔwĩ(ː),ʃa(ː)*
- 18 おまえ(たち) *ʔura(kja)/ʔja(kja)*

[柴田, 1982]

奄美大島言語地図 等語線 (isogloss)

図3 南北の対立



スナップショット
系統上の分岐?
中心からの波

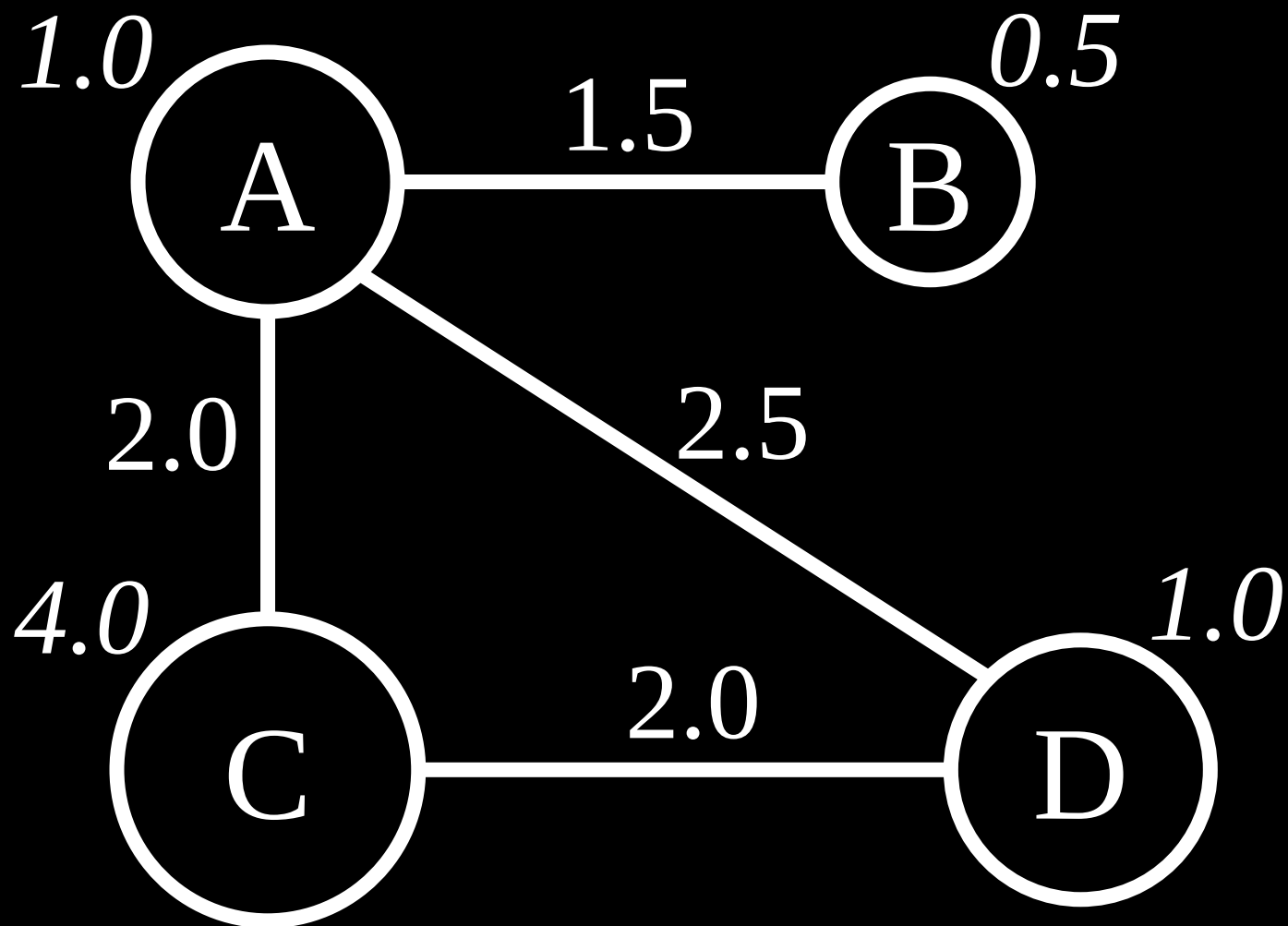
- 01 袖 *sudĩ*, etc./*sudĩ*, etc. (語末母音の長音化)
- 02 心 *kuho(ː)ro/kuːro* (語中のh音)
- 03 酒 *sĩhẽ, sɛxɛ/sĩ, sɛː* (語中のh音)
- 04 膝 *tĩbu(ː)/tsĩbuʃi, tubuʃi* (語頭子音)
- 05 とんぼ *j(ih)e:da/he:da* (語頭音節)
- 06 猫 *mjaɣ/maja* (縮約)
- 07 かまきり *ĩjito/ĩjato* (第2音節の母音)
- 08 里芋 *ʔumu/ʔuɣ, maɣ*
- 09 わたしは *wagna/waɣ, waɣnja*
- 10 いらっしゃる *ʔumorjuŋoː/ʔimorjuŋoː*
- 11 かたつむり *tĩdari(ɡjo)/tsĩɣnja, tsĩmmja*
- 12 なめくじ *nabekuʒiri/nabiːro(musĩ)*
- 13 蛙 *bikki/bikkja*
- 14 さつま芋 *hanu(ː)s(ɪ)/toŋ*
- 15 匂い *kada, kaza/niwẽ, etc.*
- 16 西 *ʔiri, ʔirui/nĩji*
- 17 上座, 下座 *wa:ra, ʃa:ra/ʔwi(ː), ʃa(ː)*
- 18 おまえ(たち) *ʔura(ʃja)/ʔja(ʃja)*

[柴田, 1982]

本日のおはなし

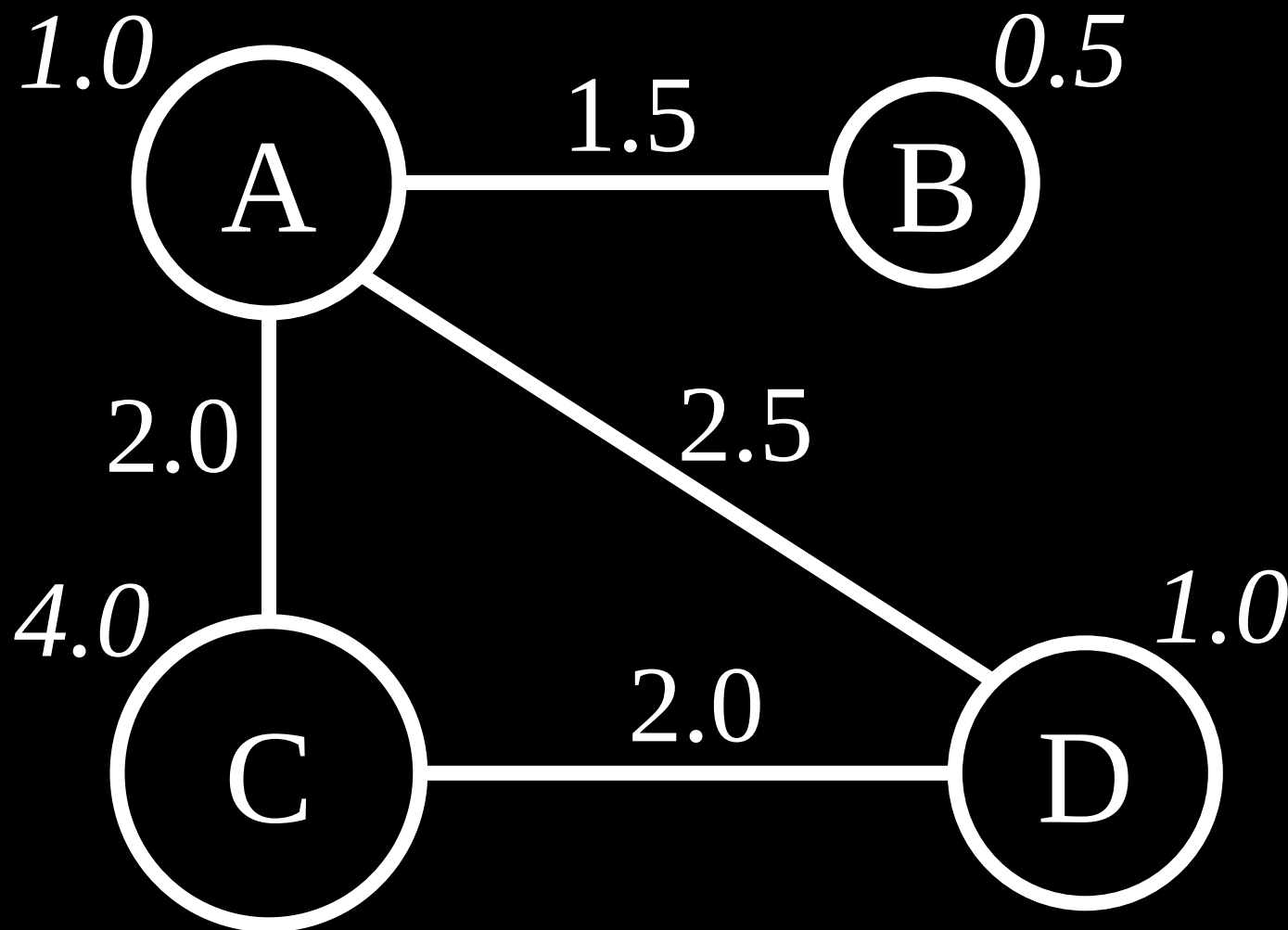
- 言語への系統モデルの適用が流行っている
- でも方言群の語彙に適用するのは無理がある
- 元データを作った人たちは何を考えていた?
- ではまったく違う接触モデルを考えてみる
- シミュレーションデータを系統モデルがどう解釈するか見てみる

接触モデル 1/3



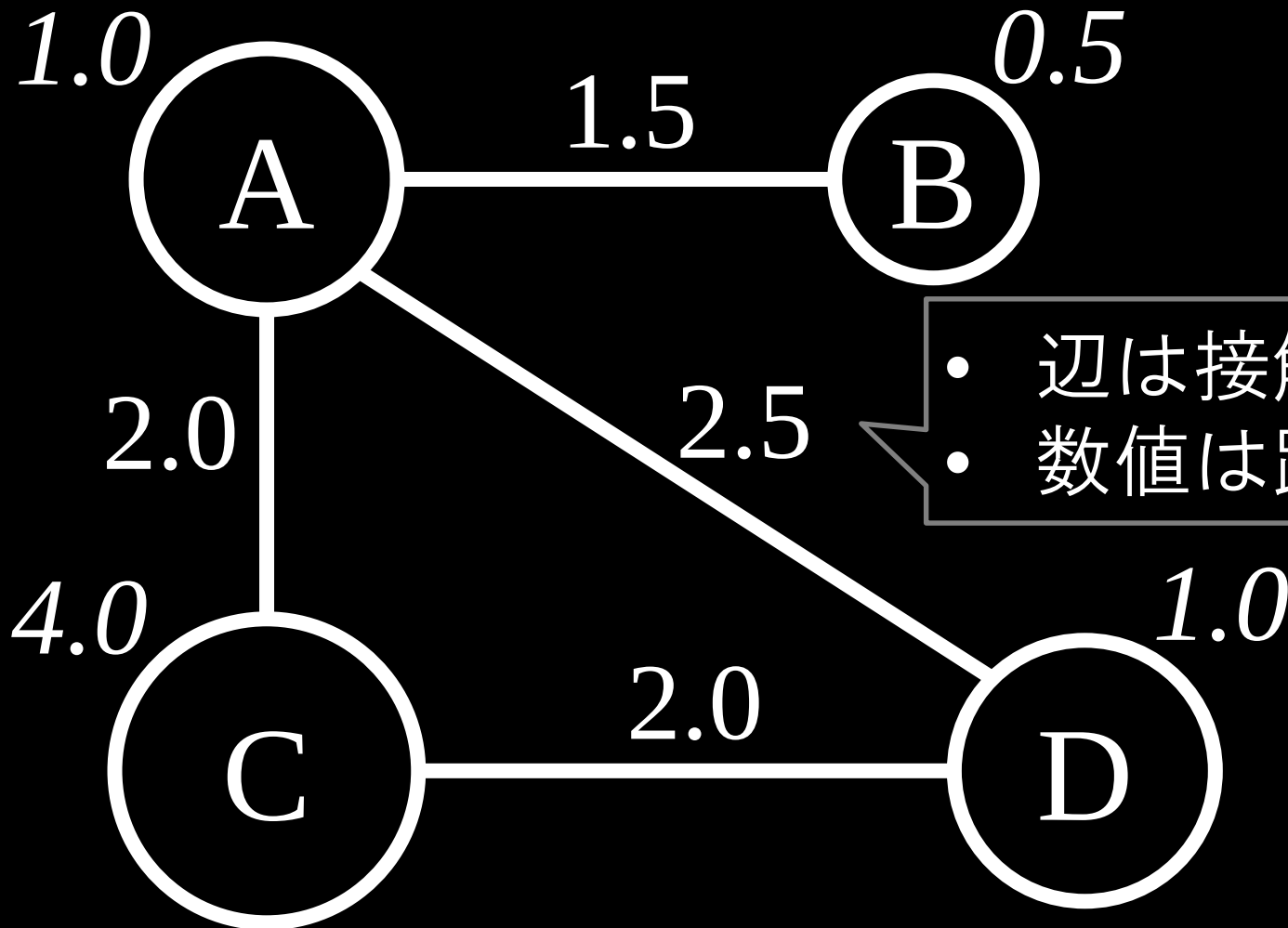
接触モデル

- 各ノードは言語
- 数値は人口



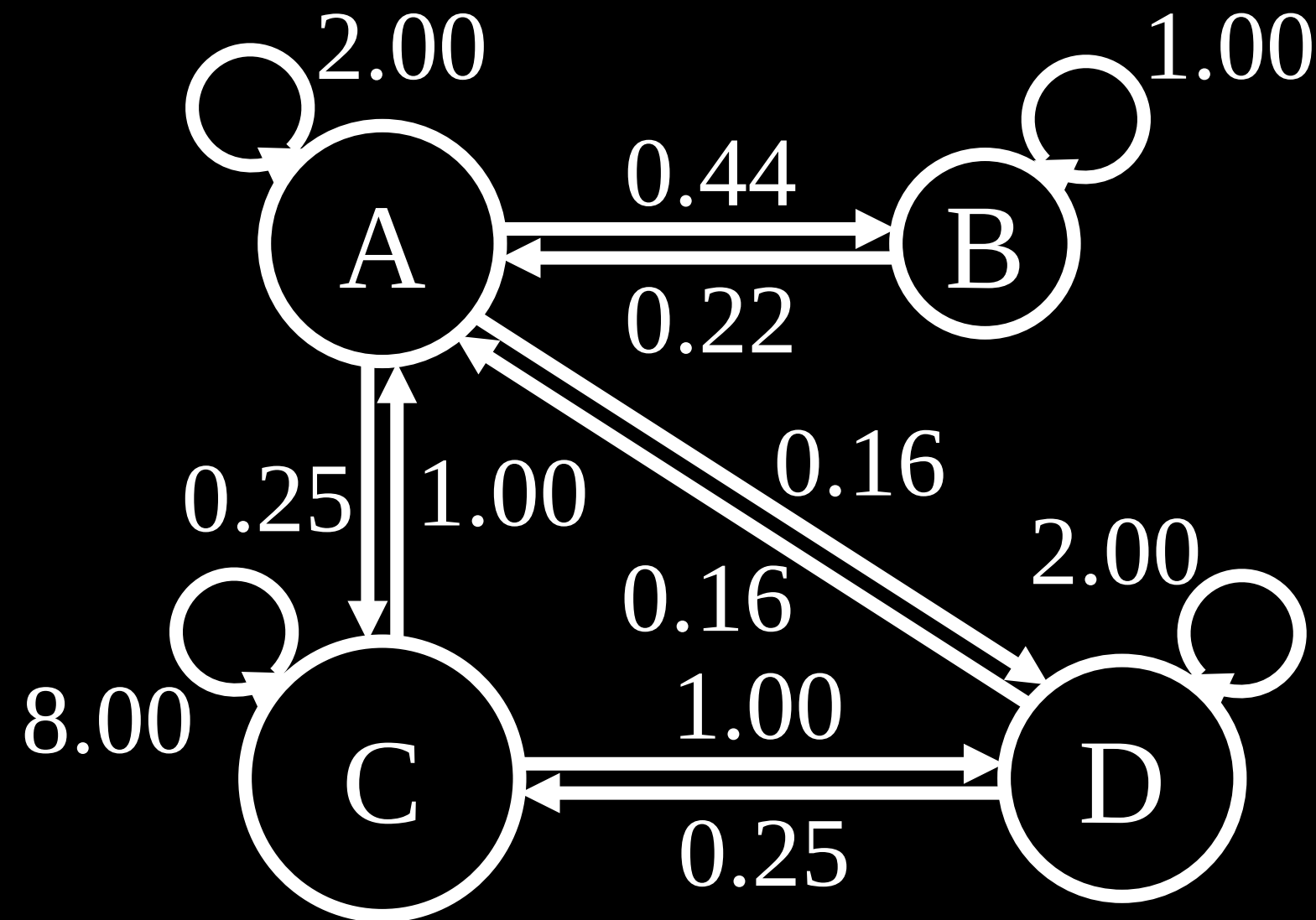
接触モデル

- 各ノードは言語
- 数値は人口

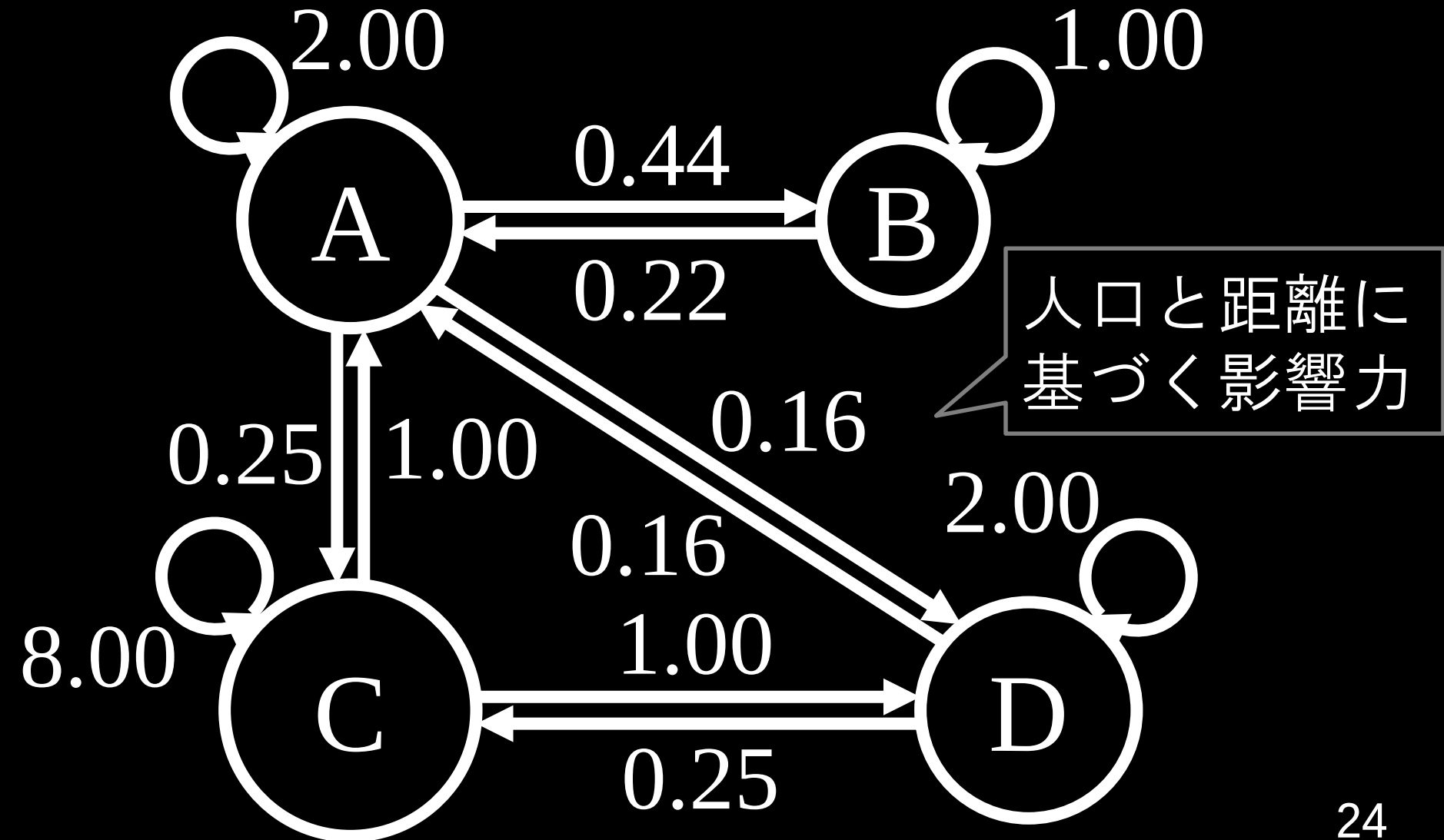


- 辺は接触関係
- 数値は距離

接触モデル 2/3

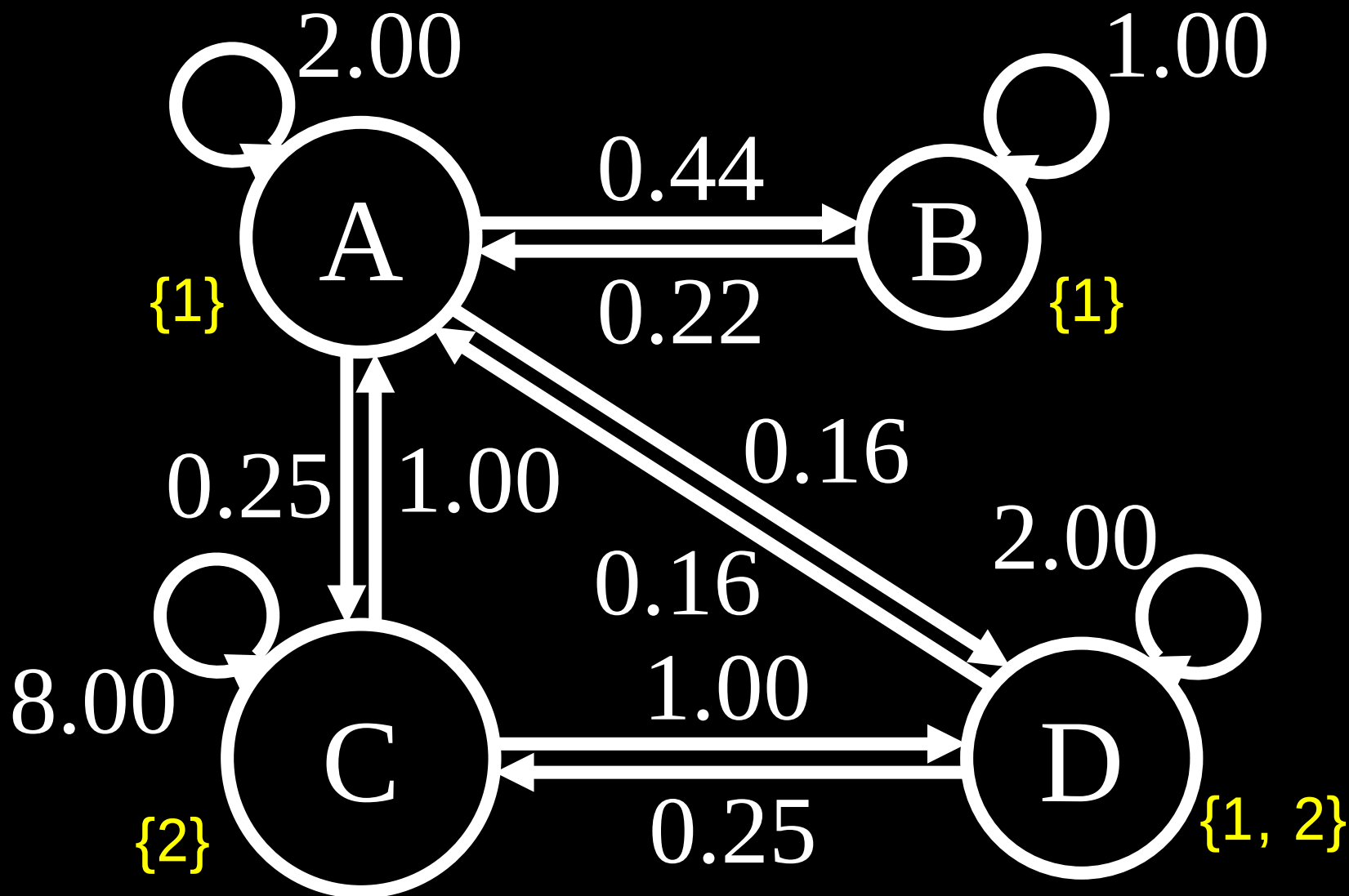


接触モデル 2/3



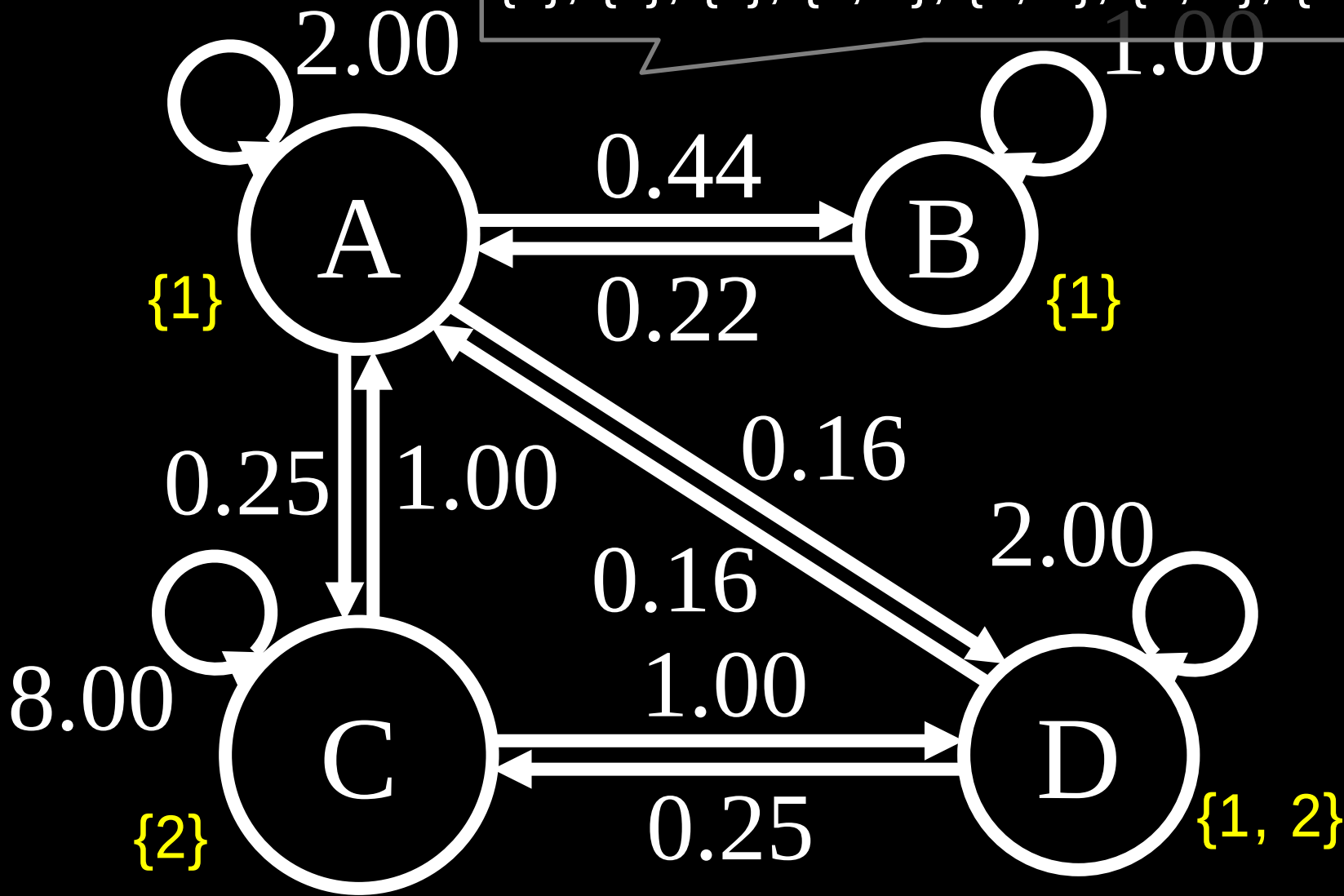
時刻 $t - 1$

接触モデル 3/3



時刻 $t - 1$

Aが時刻 t で取り得る語集合:
隣接ノードで用いられている語+新語
 $\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$



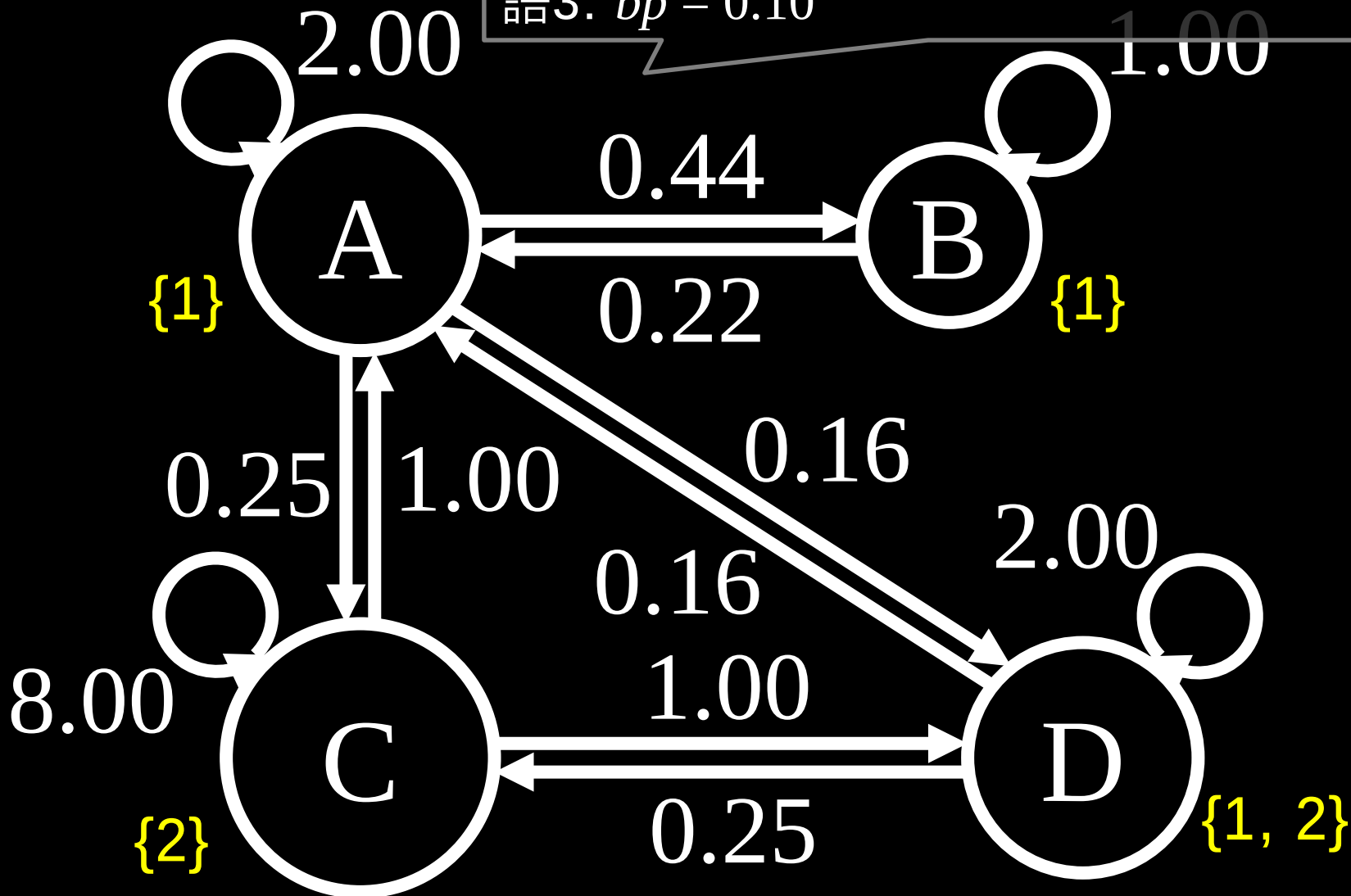
Aについて各語が持つスコア:

語1: $2.00 + 0.22 + 0.16 / 2 = 2.30$

語2: $1.00 + 0.16 / 2 = 1.08$

語3: $bp = 0.10$

時刻 $t - 1$



時刻 $t - 1$

Aについて語集合が持つスコア:

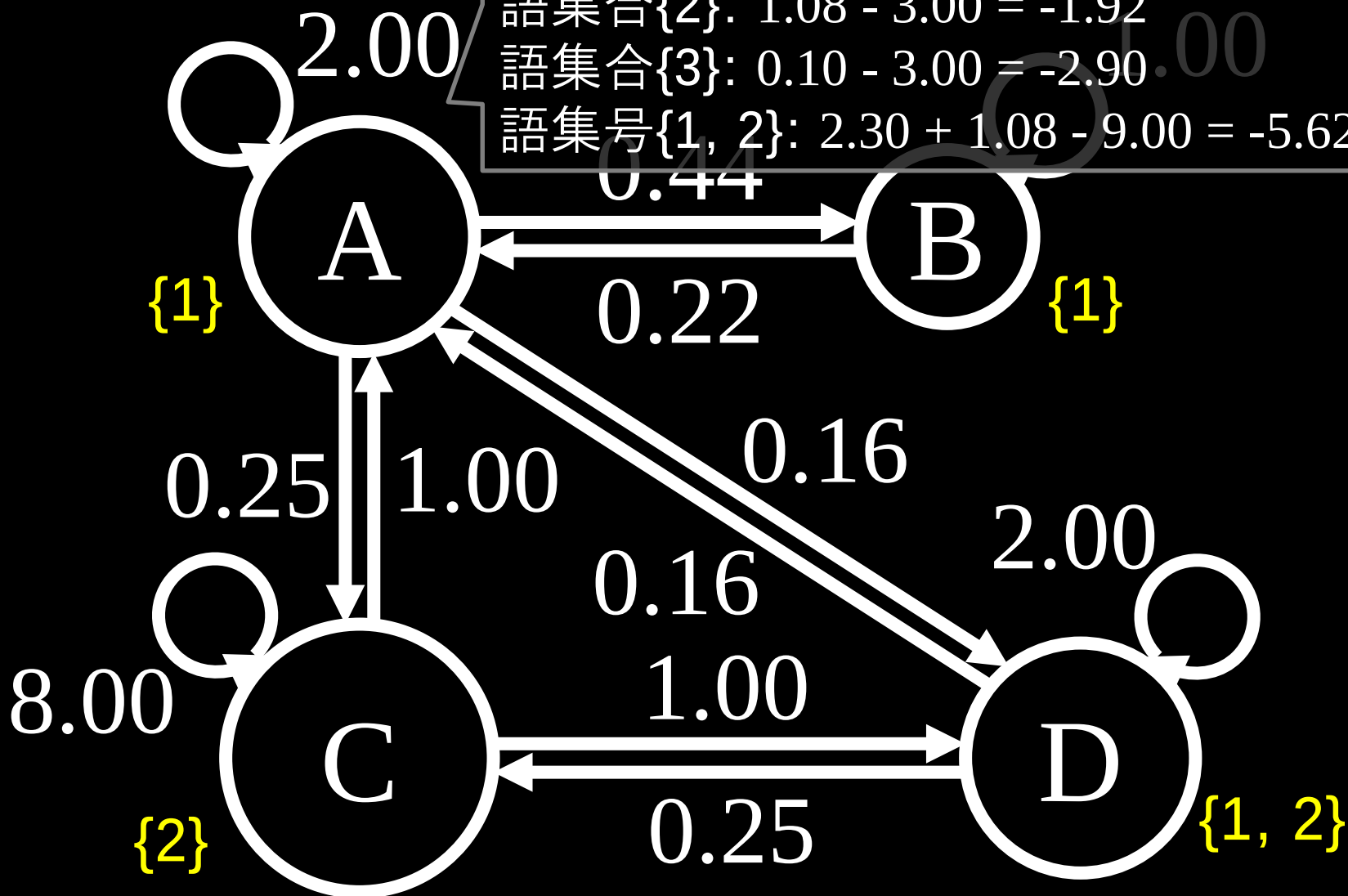
接語のスコアの総和 - 語数ペナルティ

語集合{1}: $2.30 - 3.00 = -0.70$

語集合{2}: $1.08 - 3.00 = -1.92$

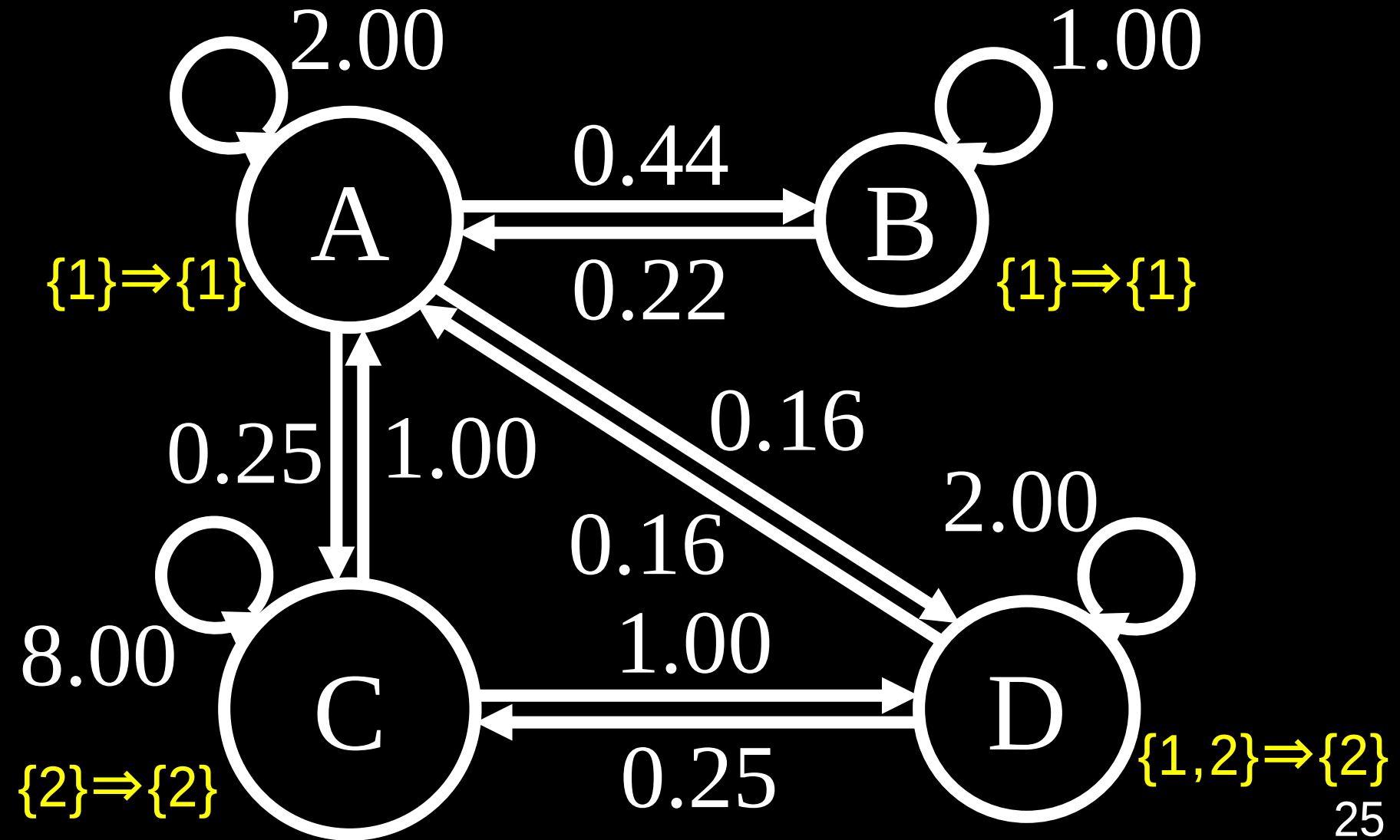
語集合{3}: $0.10 - 3.00 = -2.90$

語集号{1, 2}: $2.30 + 1.08 - 9.00 = -5.62$



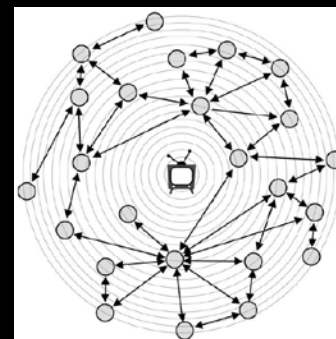
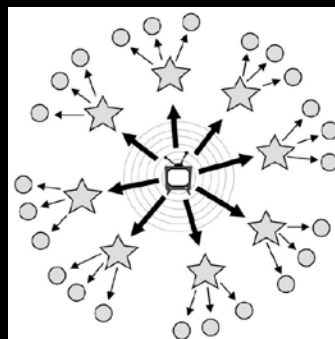
時刻 t

Aにおける各語集合のスコアをもとに
時刻 t での語集合を確率的に選択



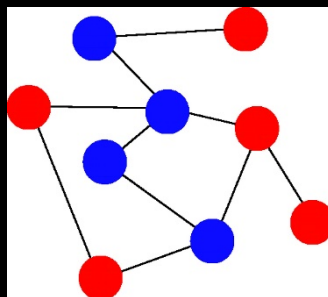
関連研究

- Word of mouth: 世論形成、マーケティング、技術革新の拡散



[Watts, 2007]

- 進化グラフ



ゲームの利得行列:

$$\begin{array}{cc} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{array} \quad [\text{Nowak+}, 2010]$$

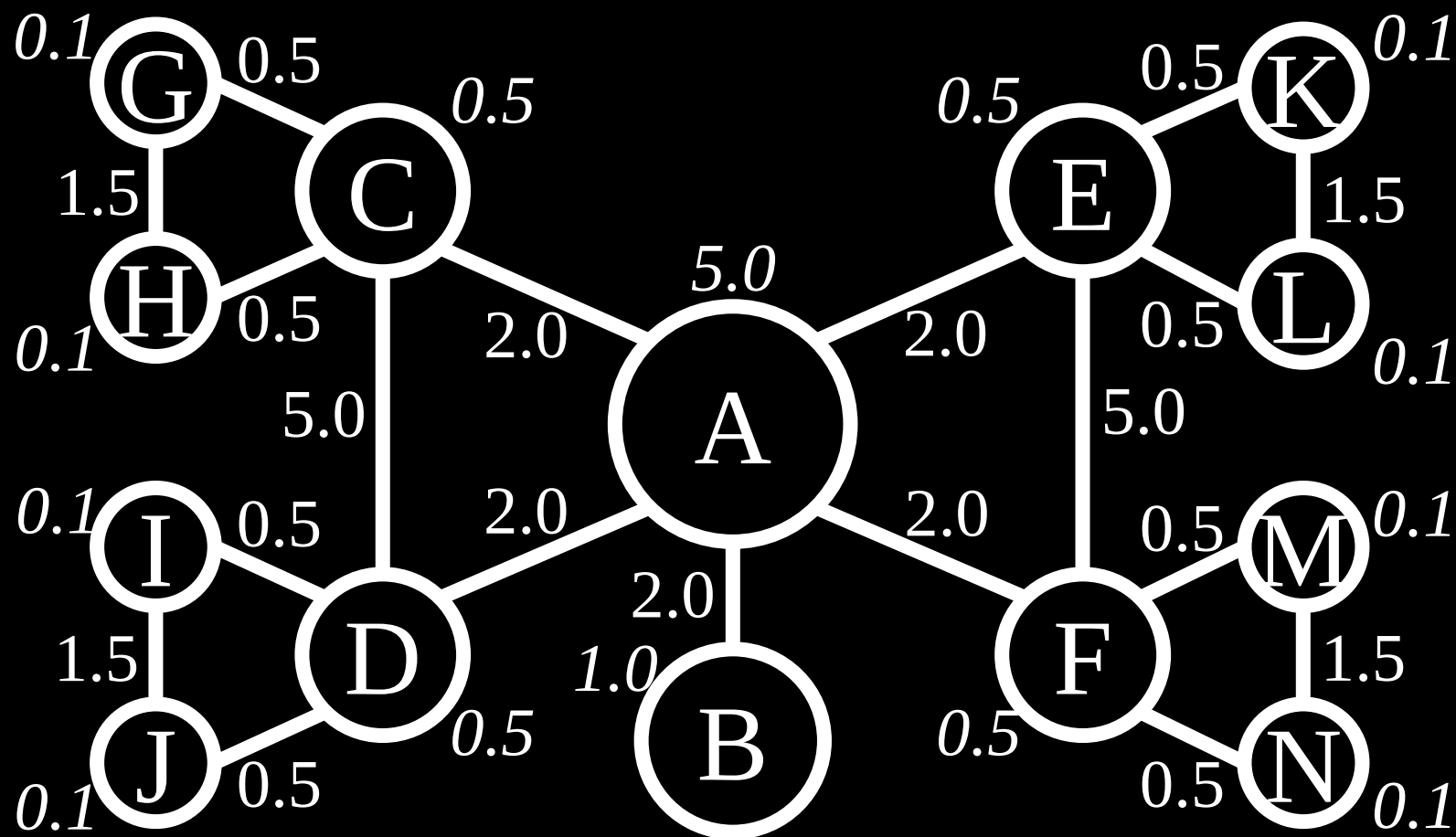
本日のおはなし

- 言語への系統モデルの適用が流行っている
- でも方言群の語彙に適用するのは無理がある
- 元データを作った人たちは何を考えていた?
- ではまったく違う接触モデルを考えてみる
- シミュレーションデータを系統モデルがどう解釈するか見てみる

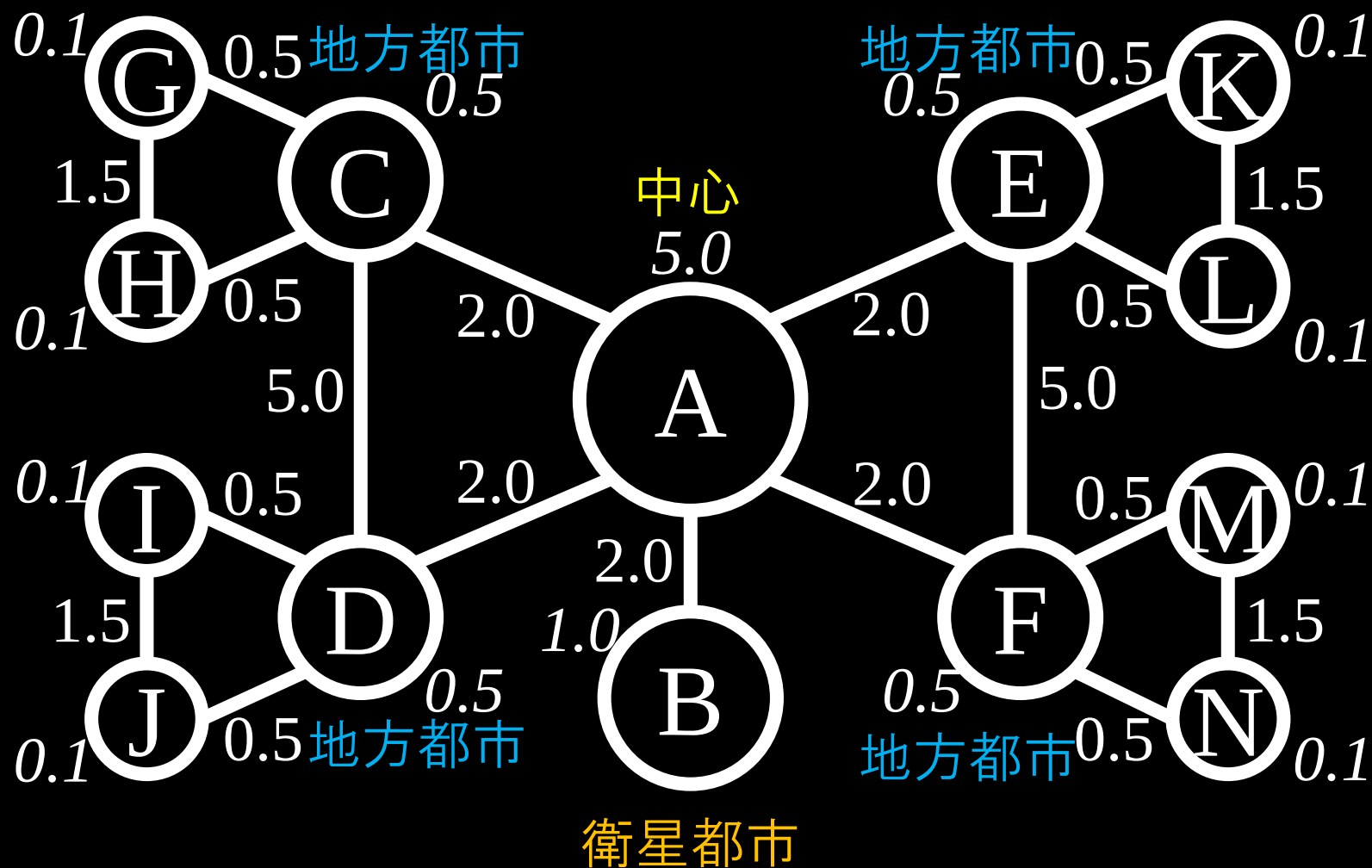
実験

- 接触モデルに様々なネットワークポロジィを与え、人工データを生成させる
 1. 星型
 2. 隘路型
 3. 殖民型
- 人工データを系統モデルに解釈させる

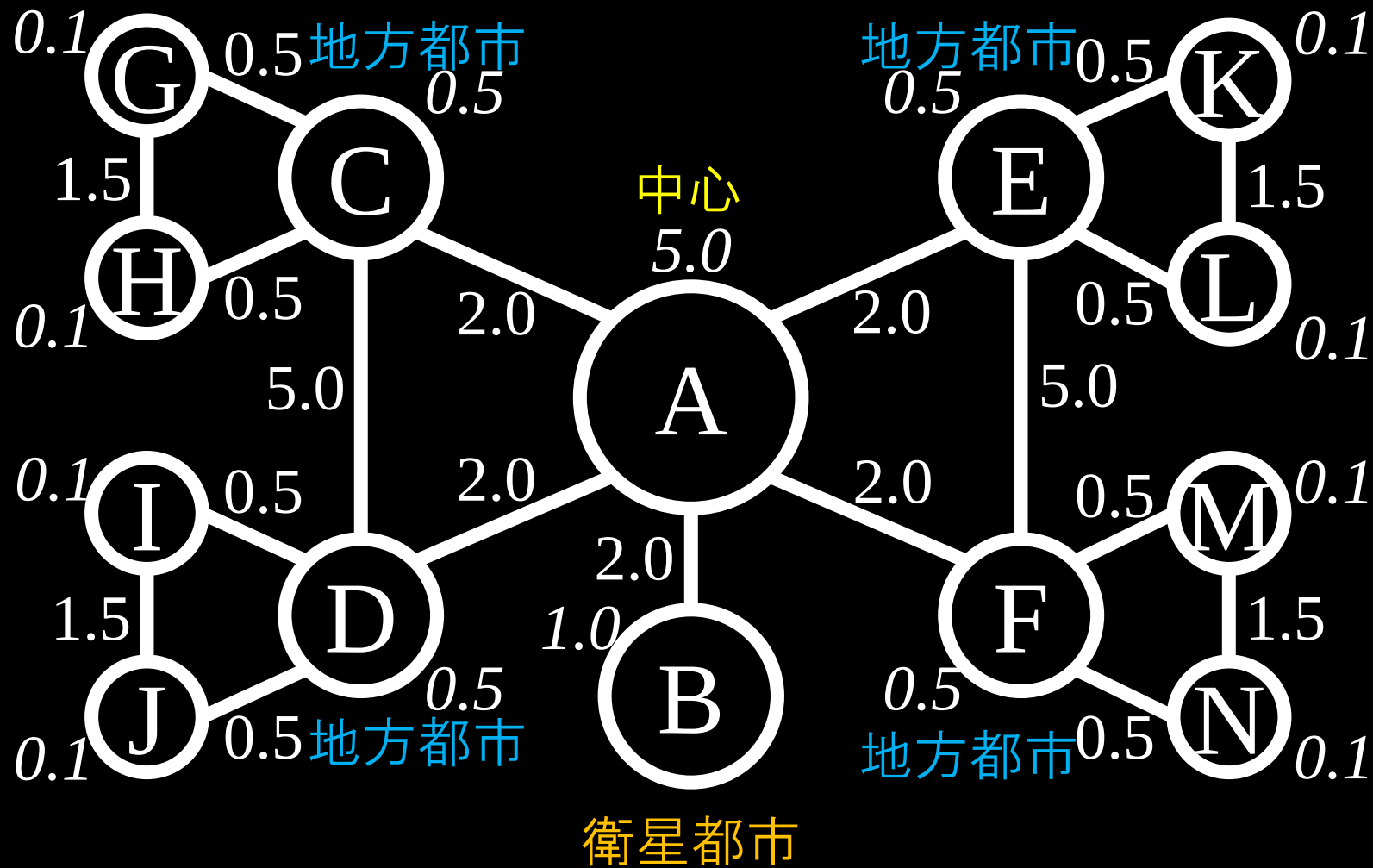
例1: 星型のトポロジー



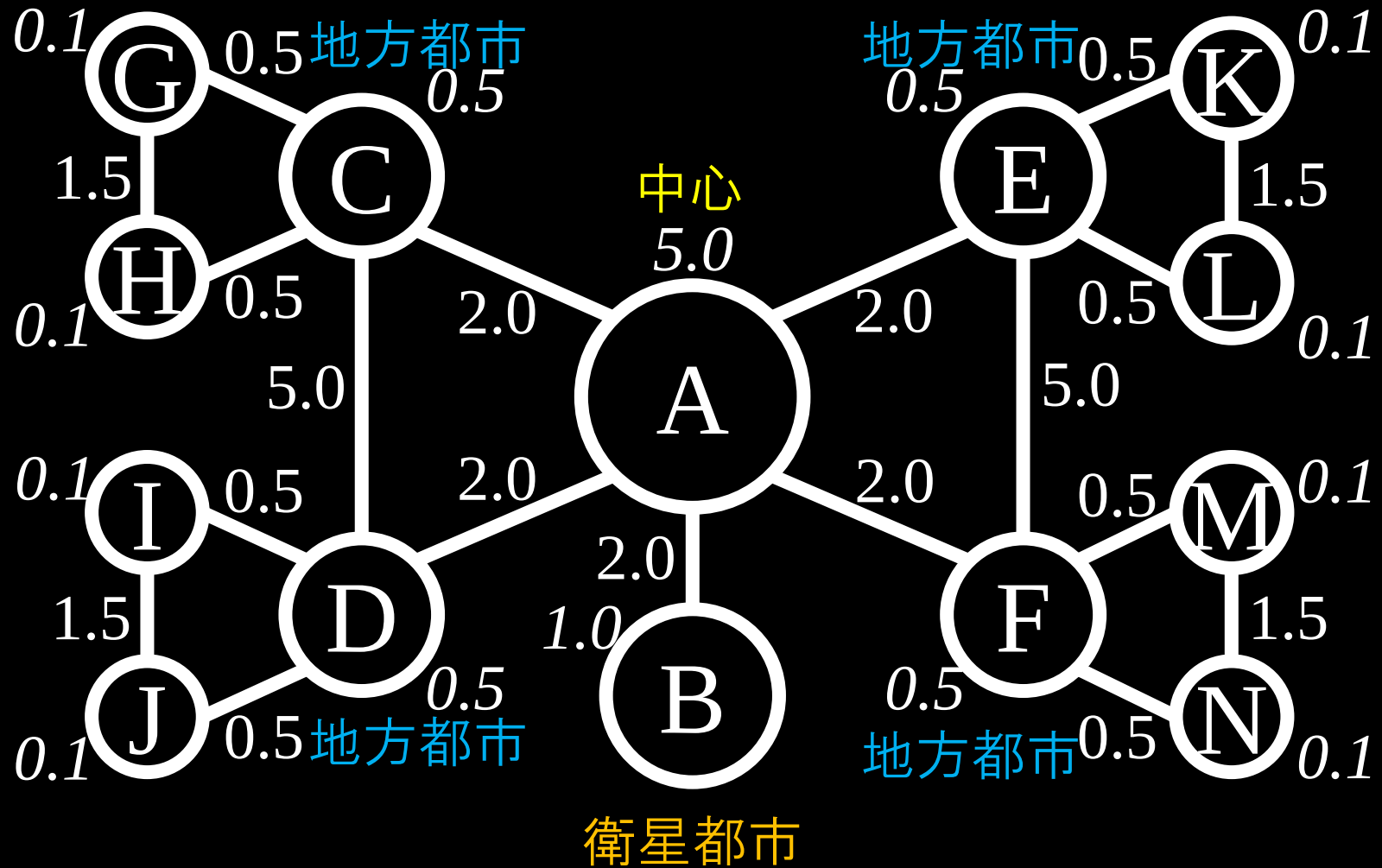
例1: 星型のトポロジー



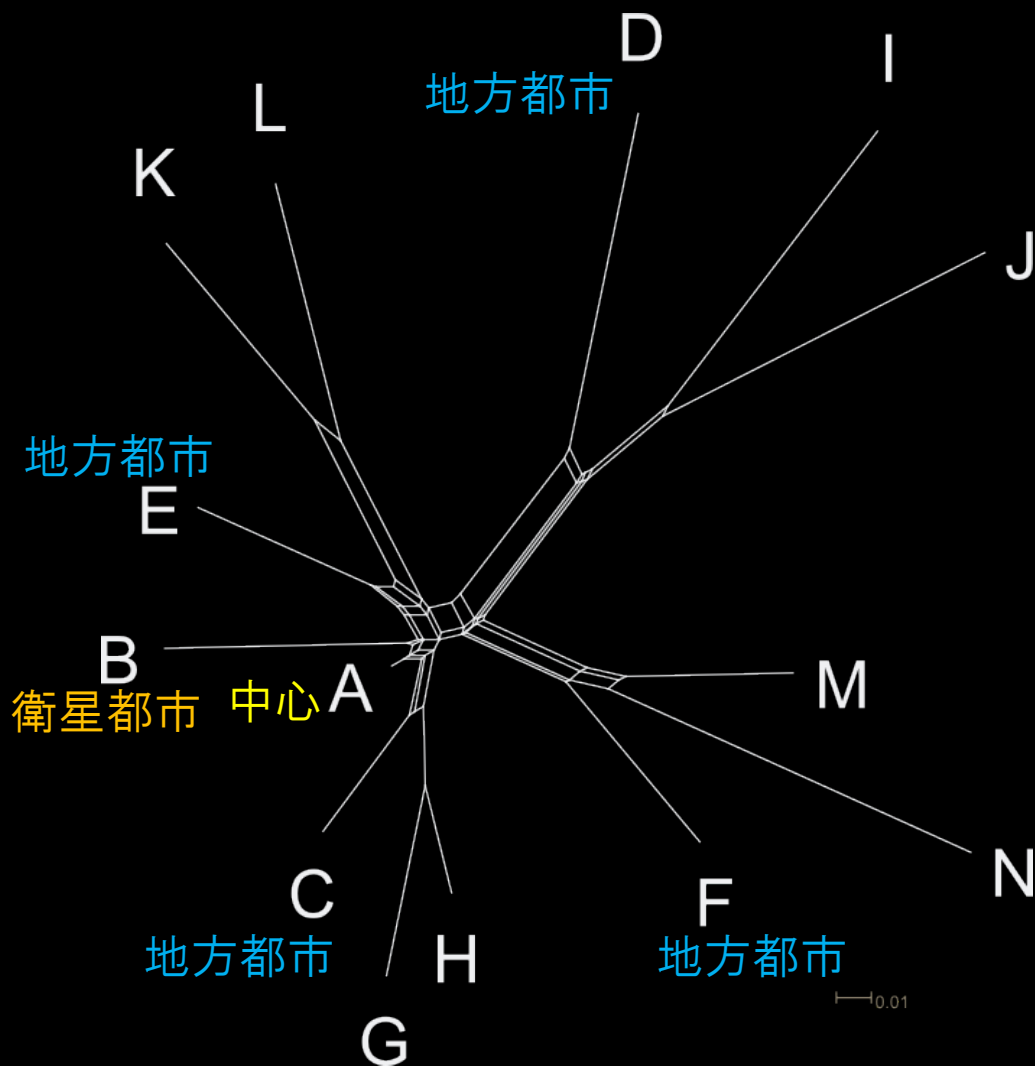
空間的に木らしい星型のトポロジー



空間的に木らしい型のトポロジに関して一定



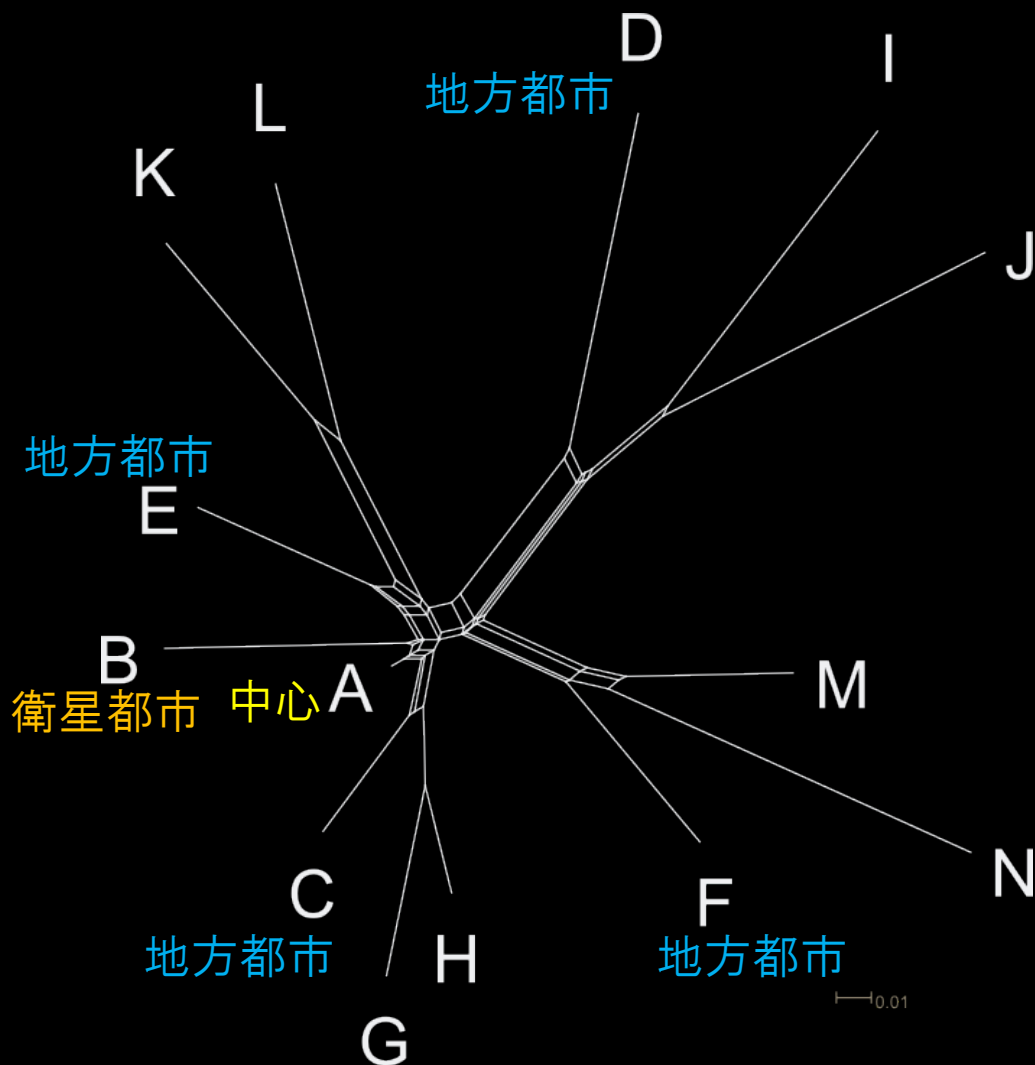
例1: 星型のNeighborNet



読み方:

- 距離に基づくクラスタリング
- 網 (平行四辺形) が矛盾を示す
- 2辺の長さ比が1に近いほど木らしくない

例1: 星型のNeighborNet

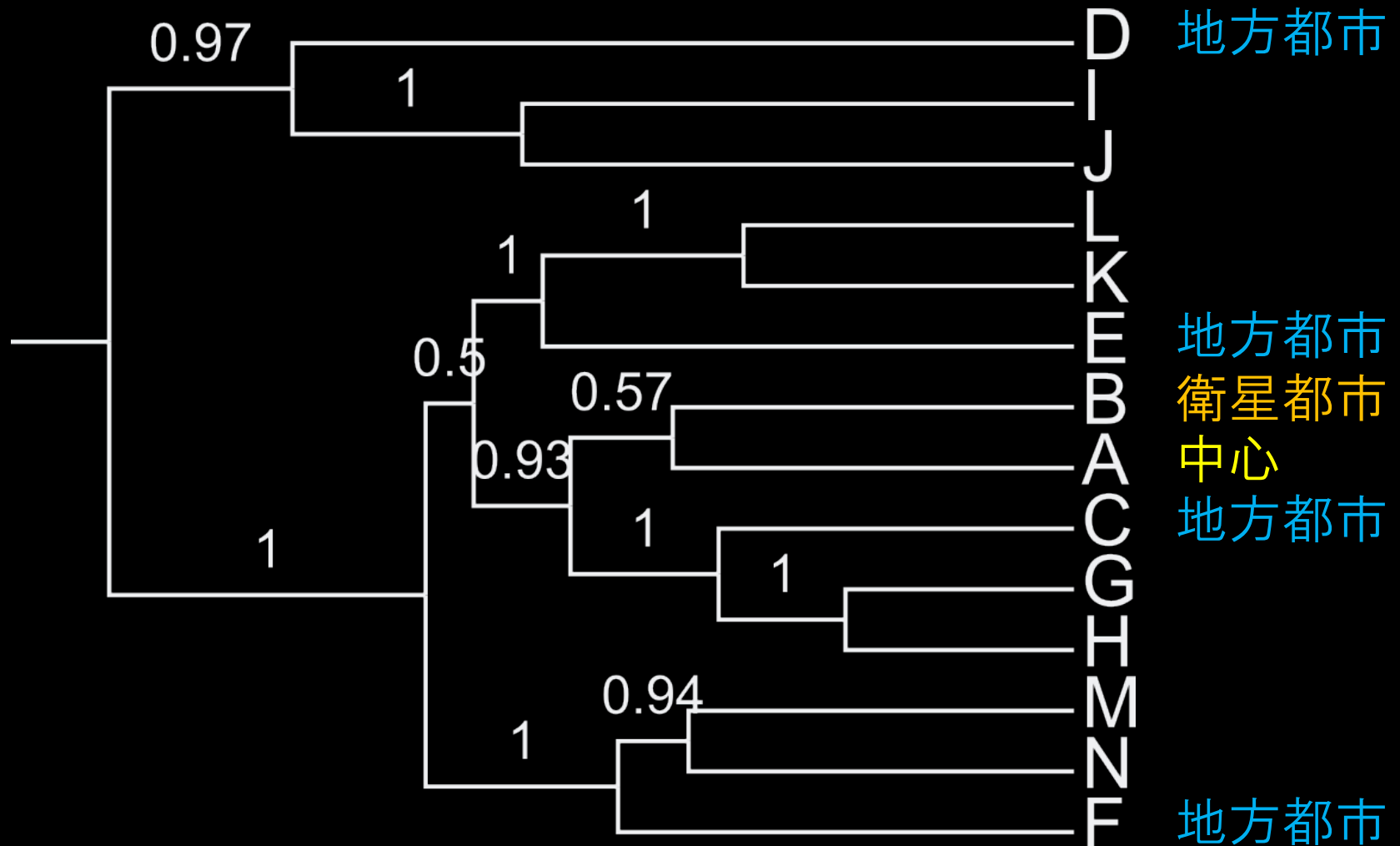


読み方:

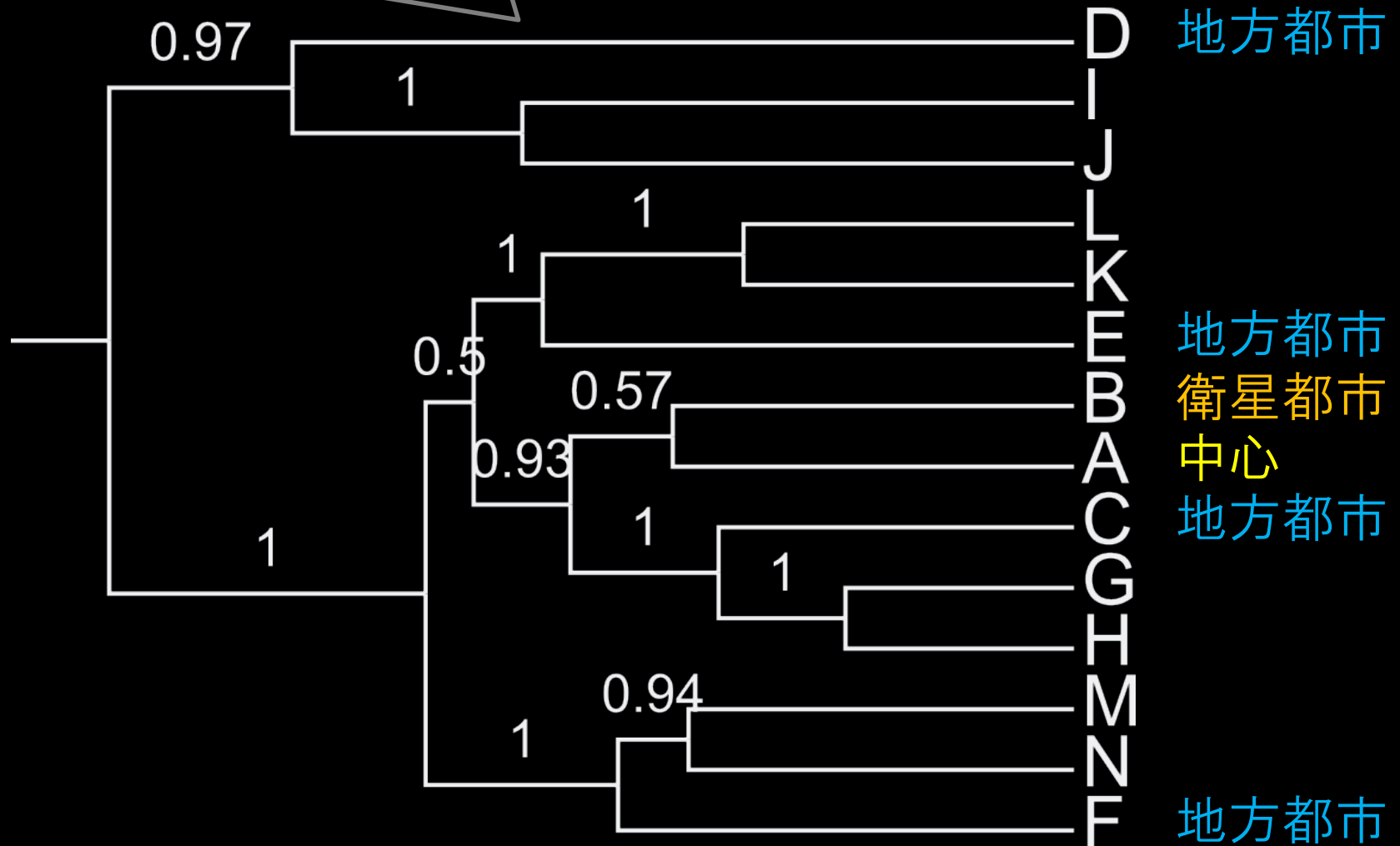
- 距離に基づくクラスタリング
- 網 (平行四辺形) が矛盾を示す
- 2辺の長さ比が1に近いほど木らしくない

結構木らしい!

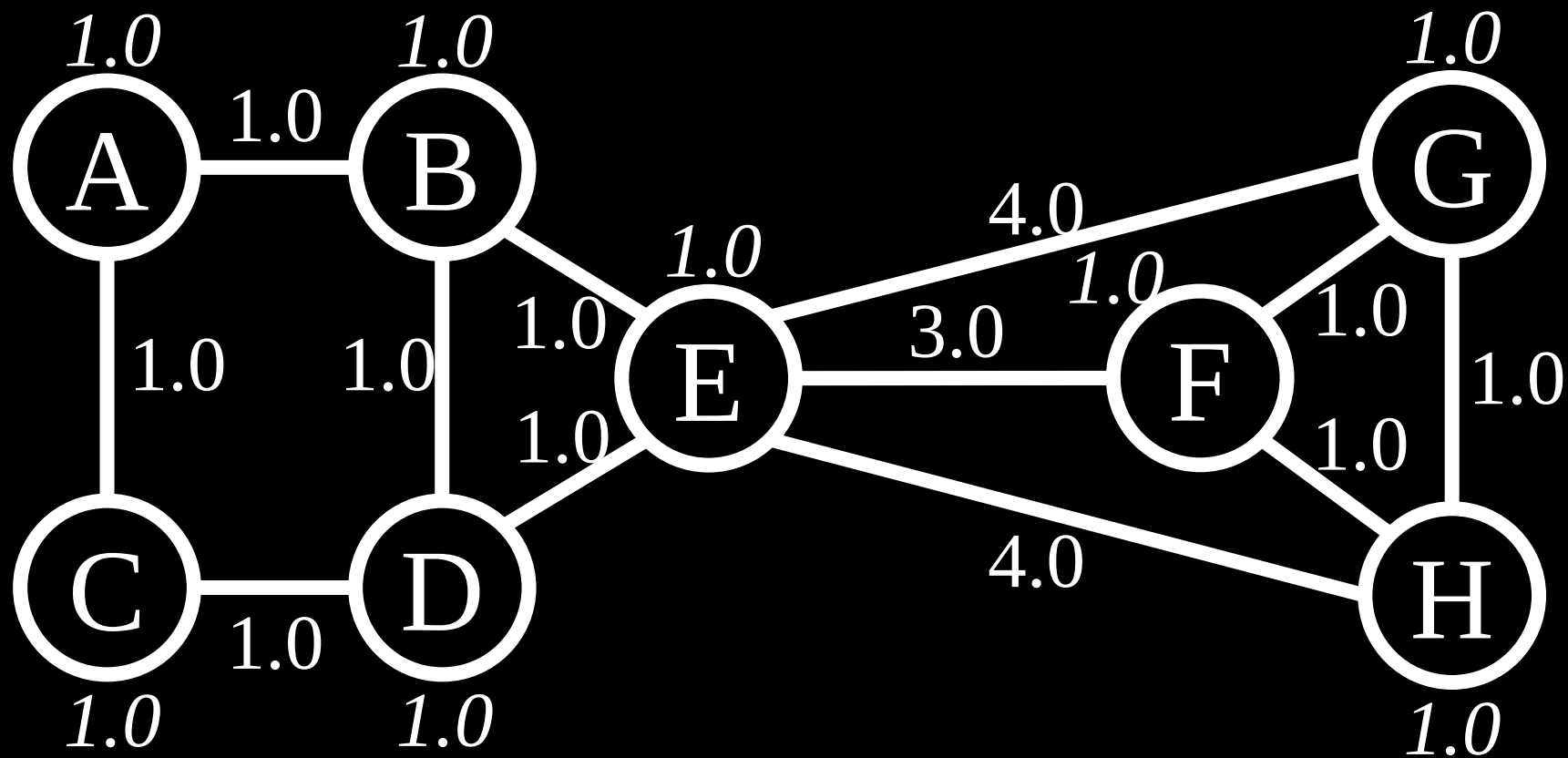
例1: 星型の系統樹



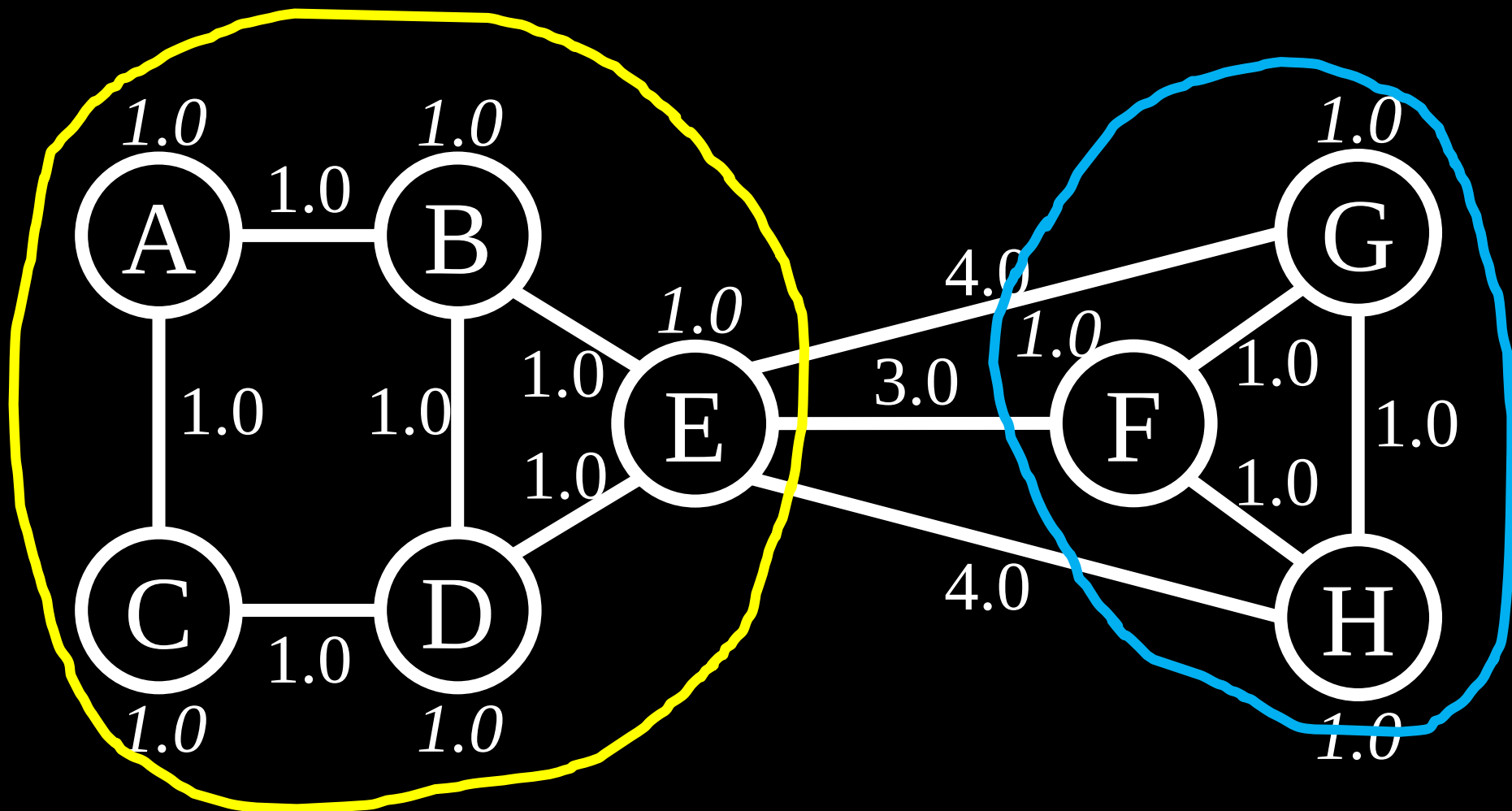
空間構造を時間構造に読み替えている星型の系統樹



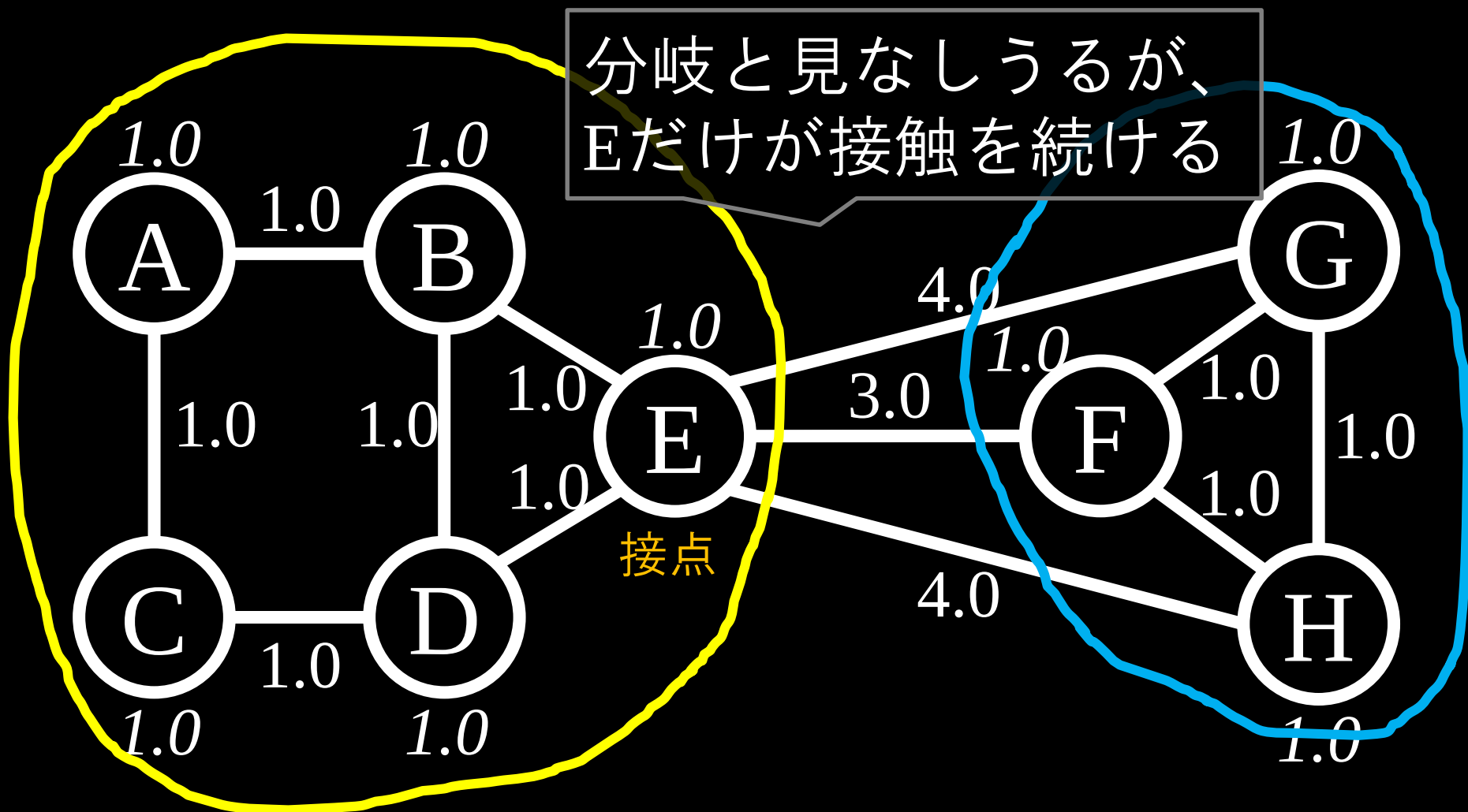
例2: 隘路型のトポロジー



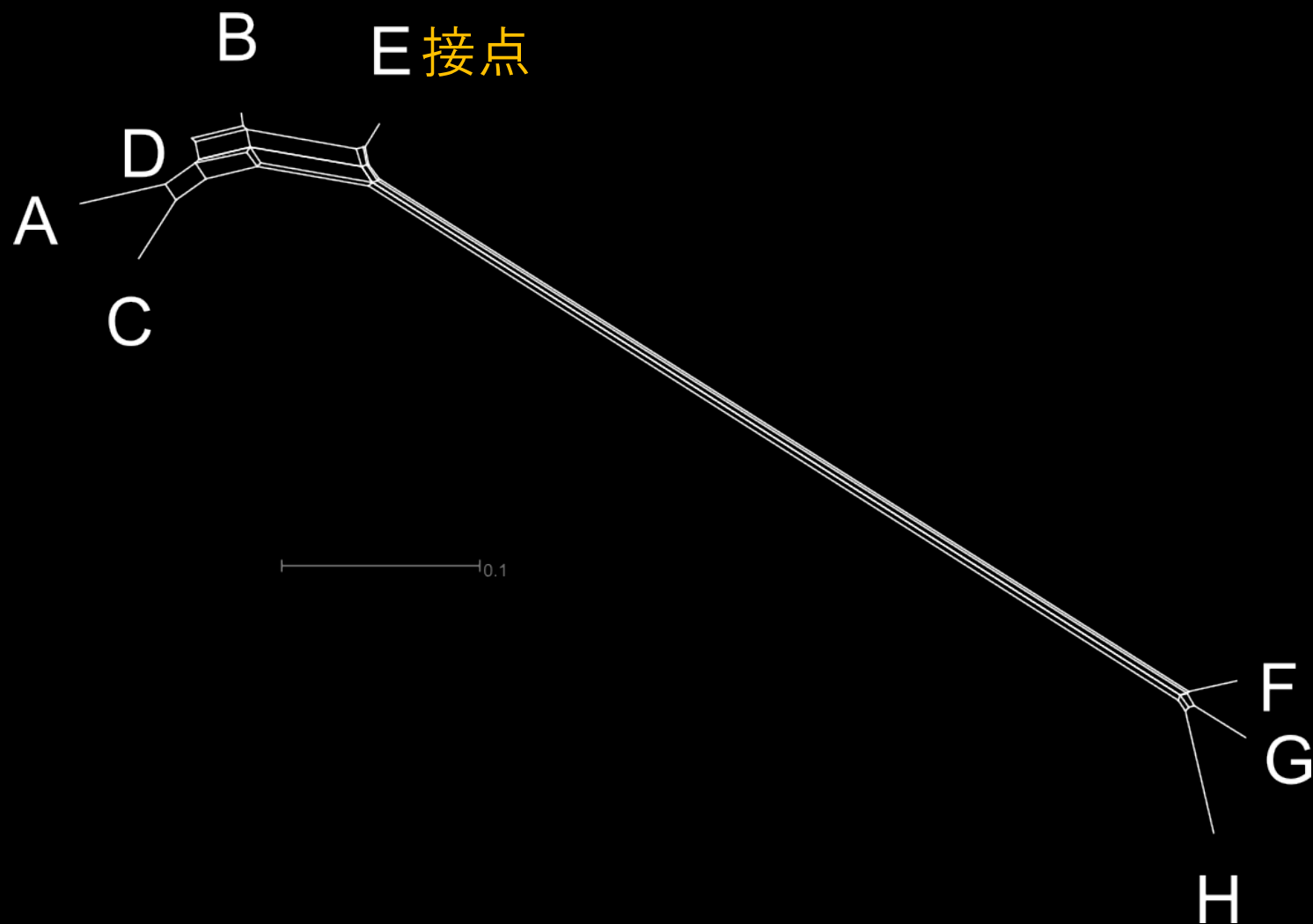
例2: 隘路型のトポロジー



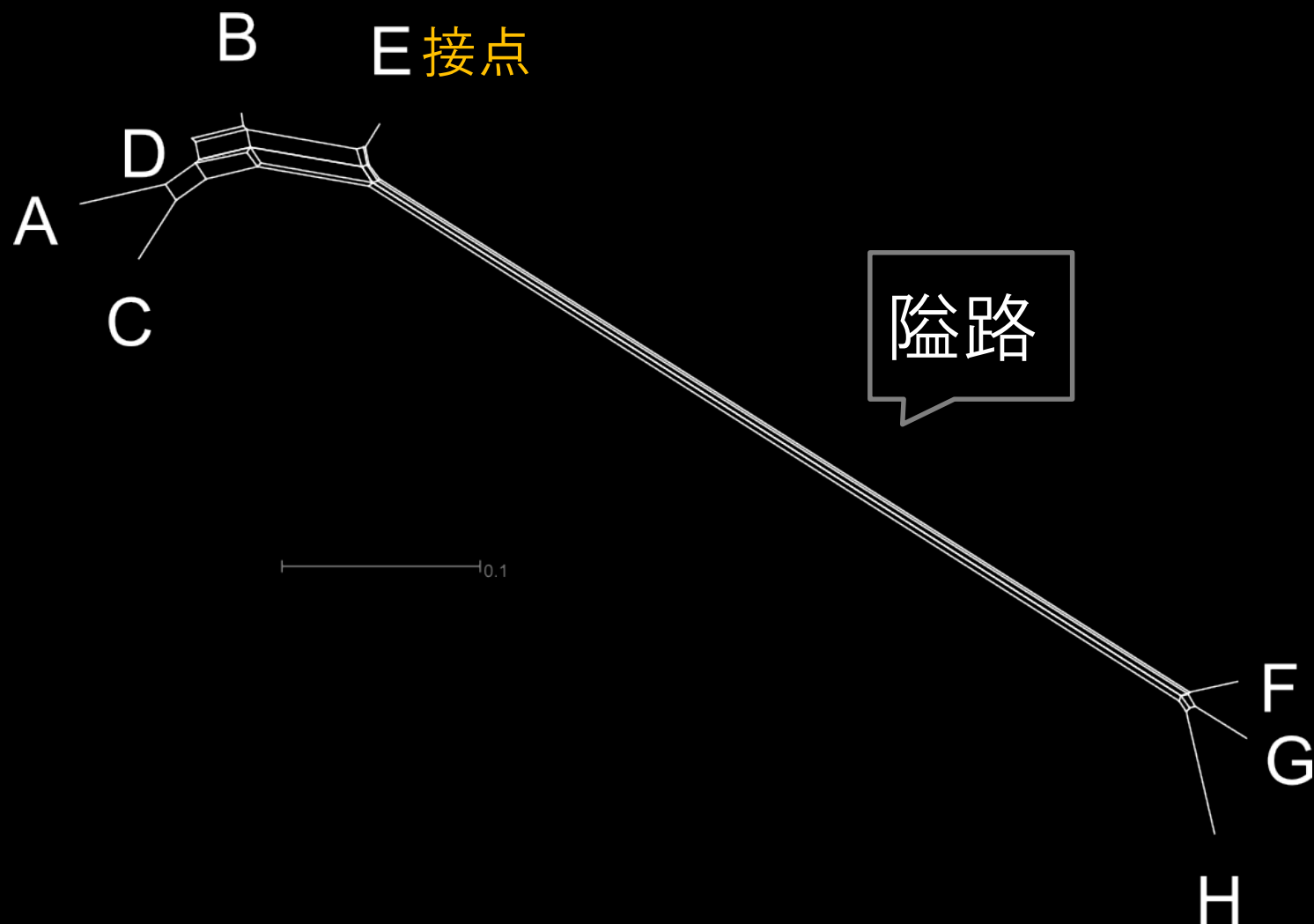
例2: 隘路型のトポロジー



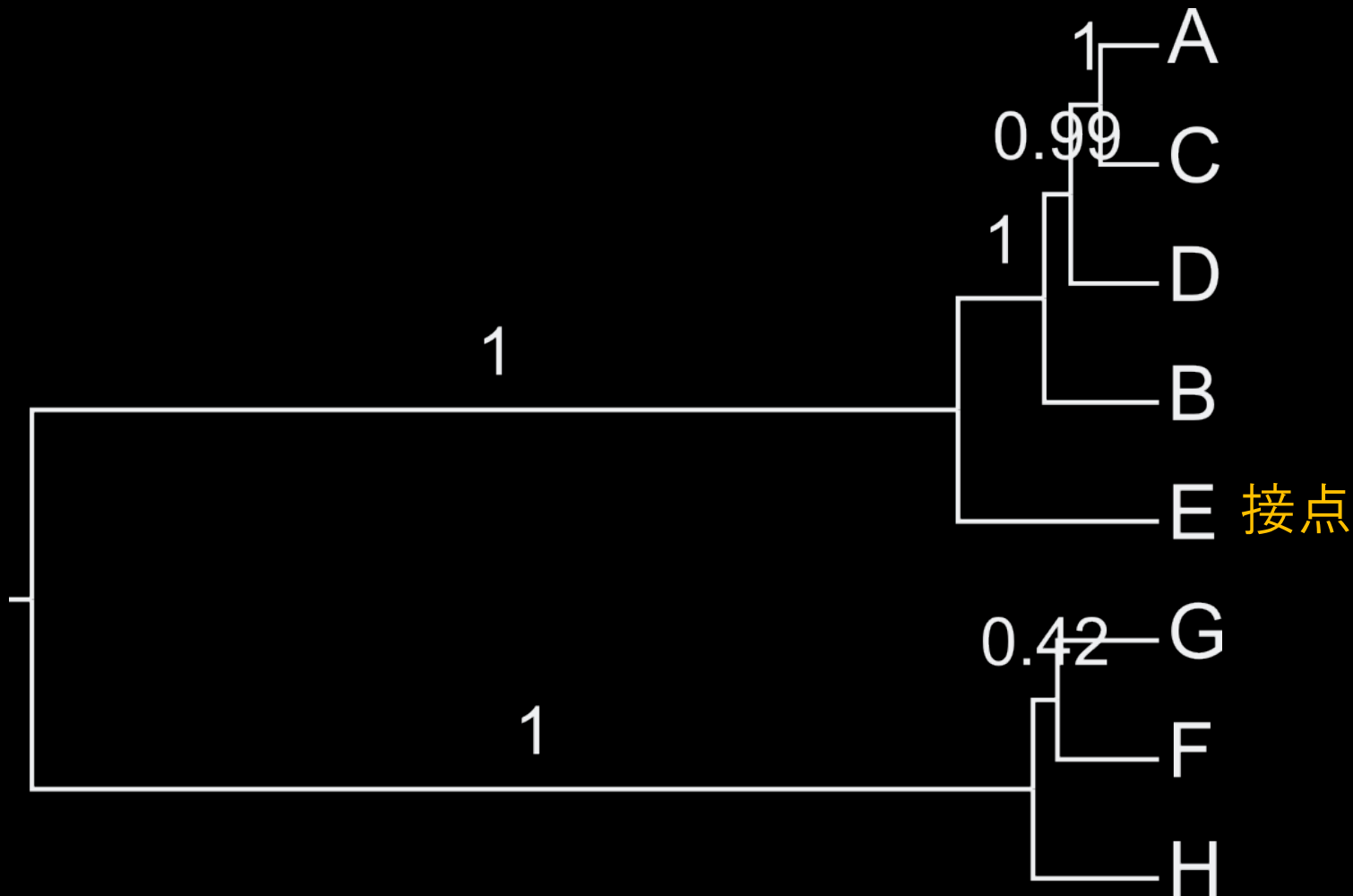
例2: 隘路型のNeighborNet



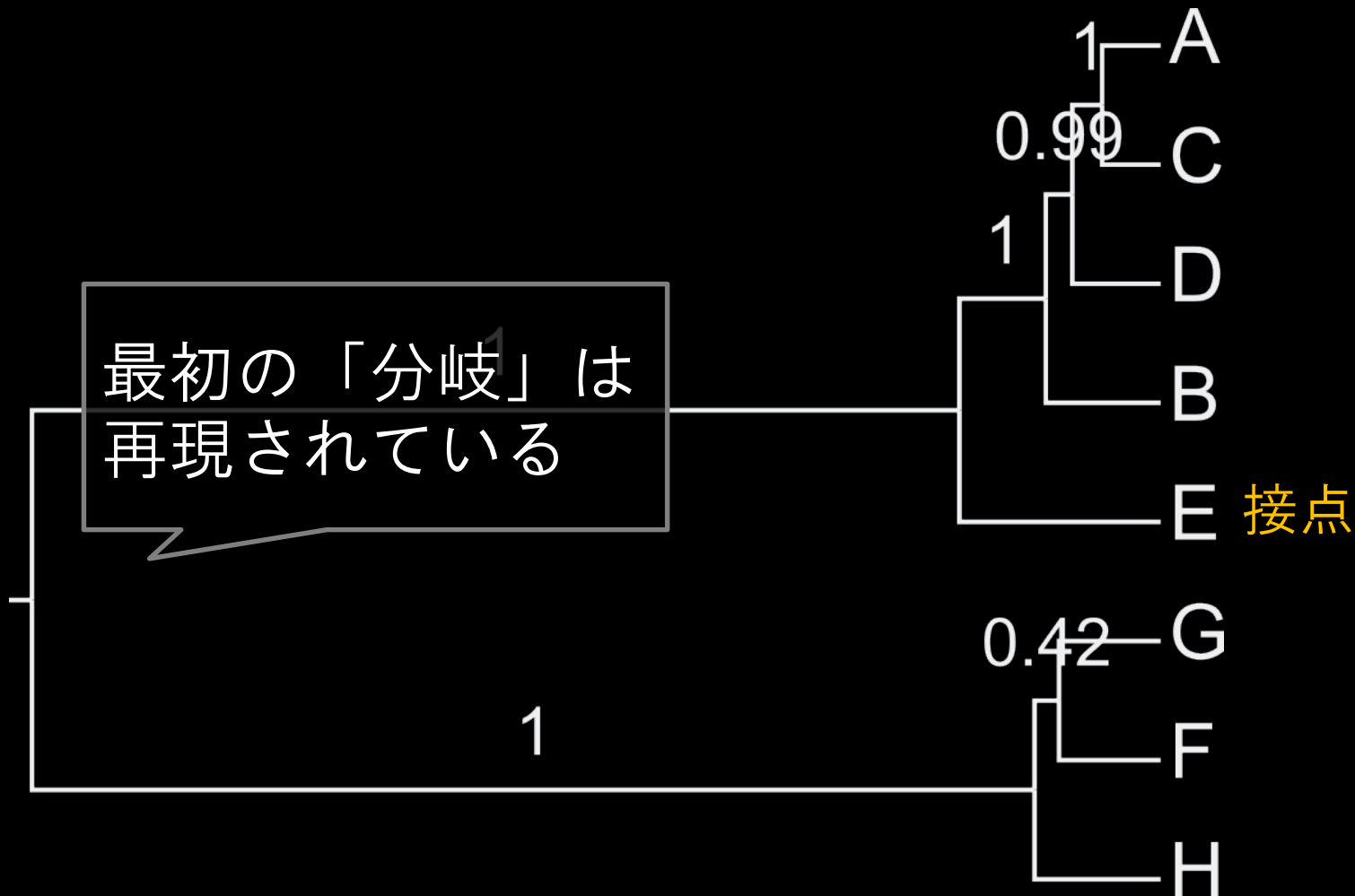
例2: 隘路型のNeighborNet



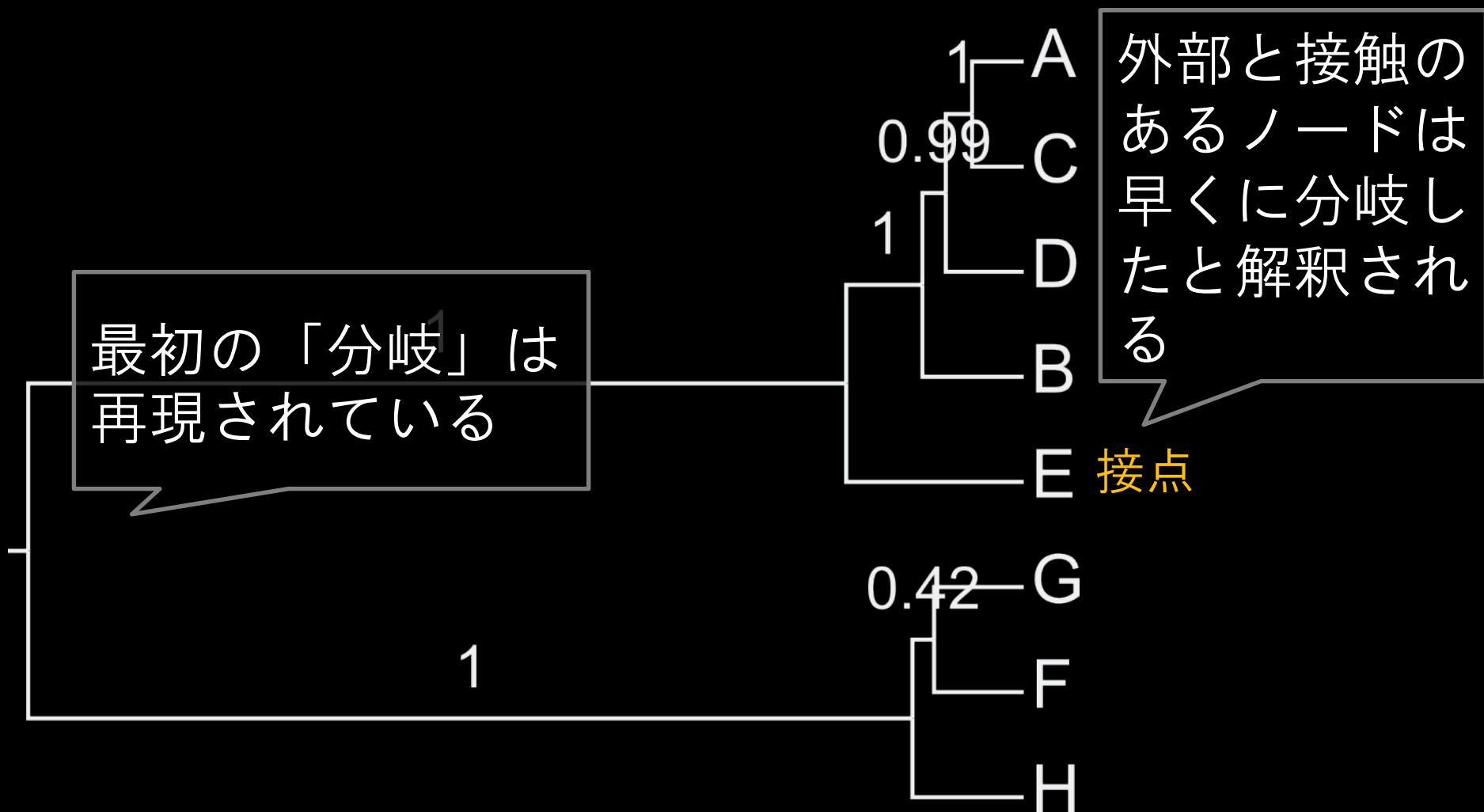
例2: 隘路型の系統樹

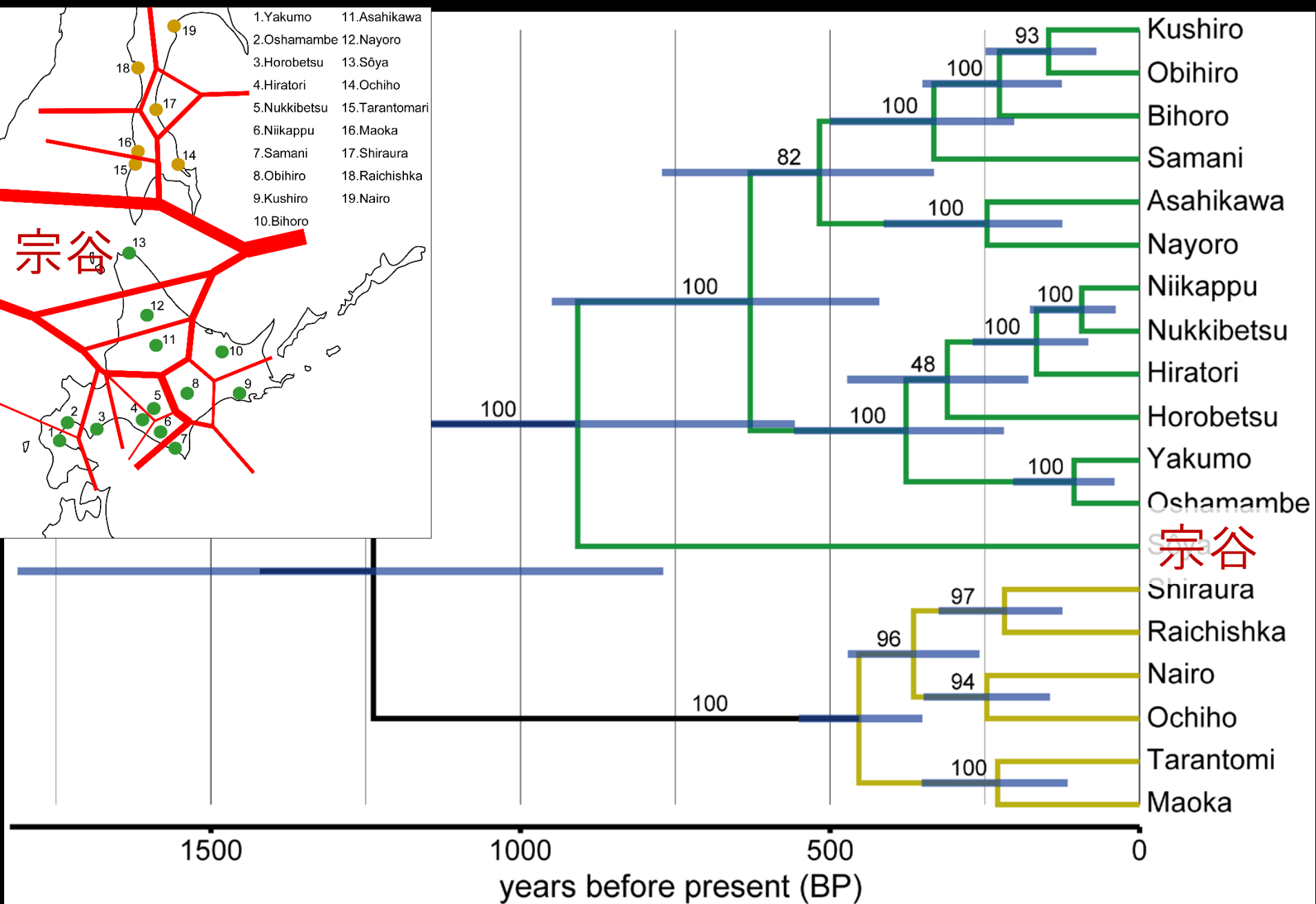


例2: 隘路型の系統樹



例2: 隘路型の系統樹



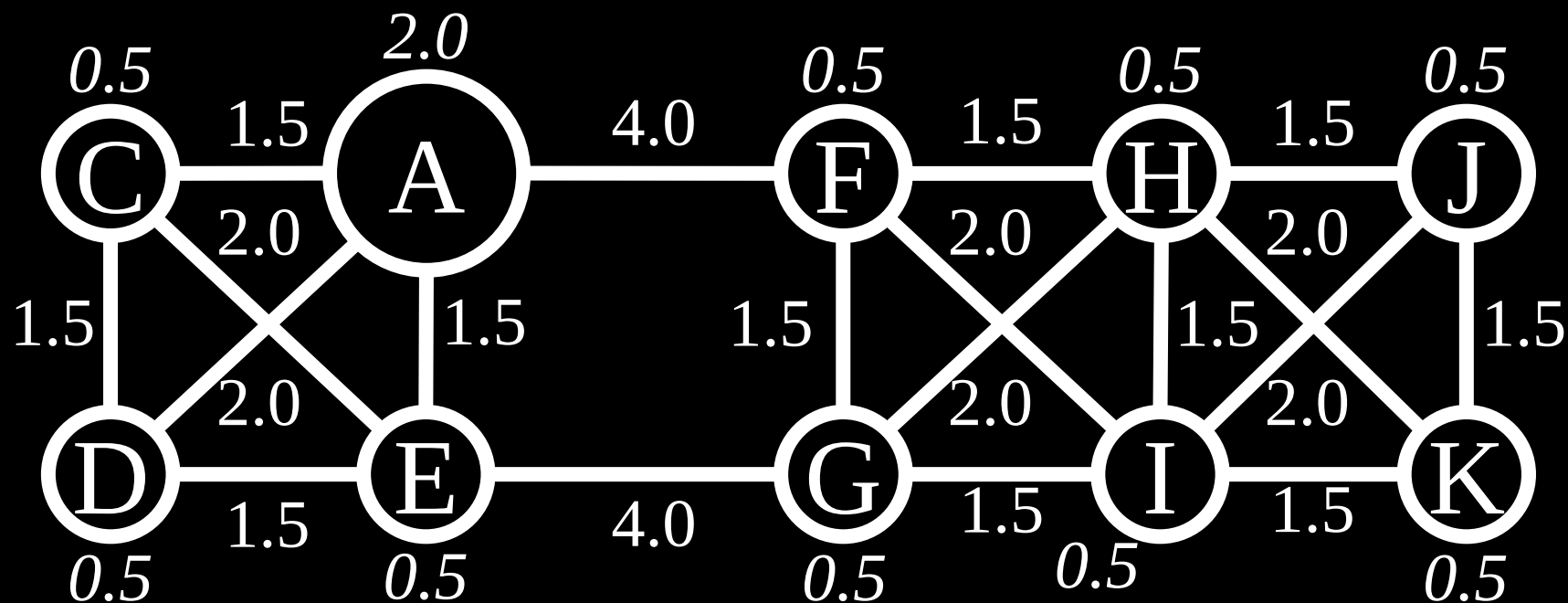




宗谷の早い分岐も
樺太との接触が原
因では？

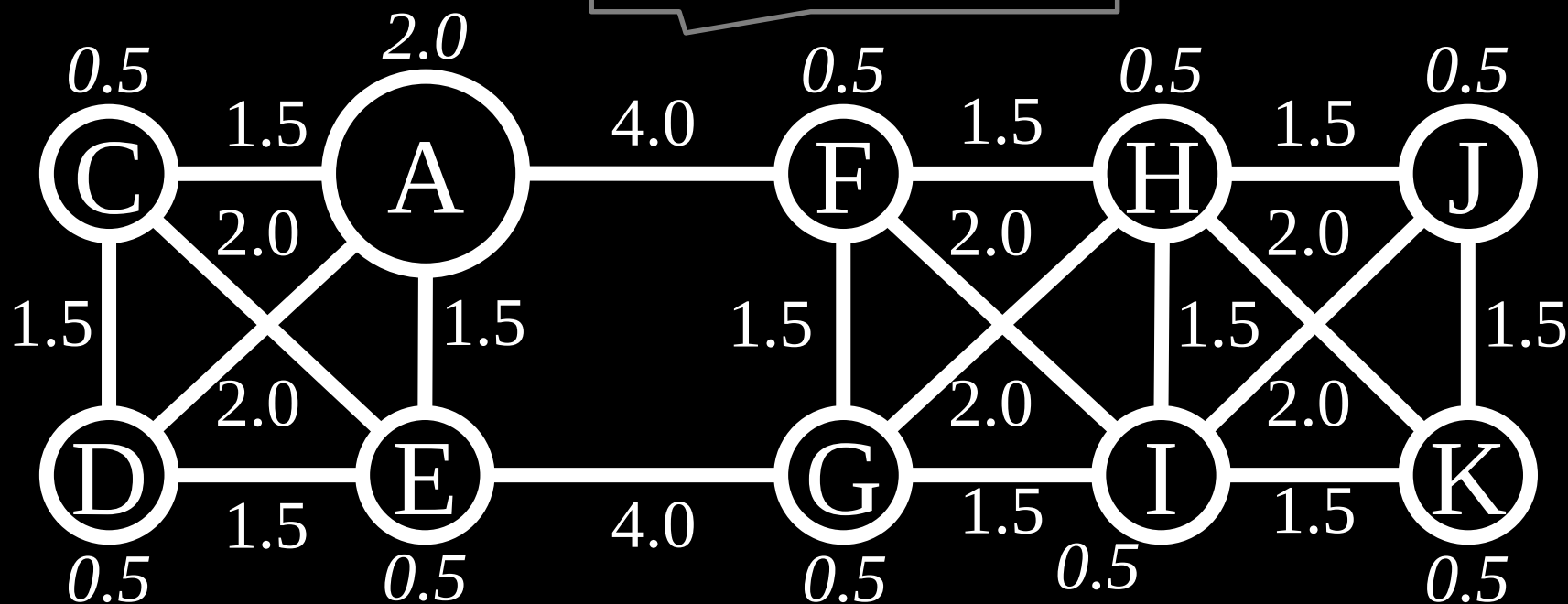


例3: 殖民型のトポロジー



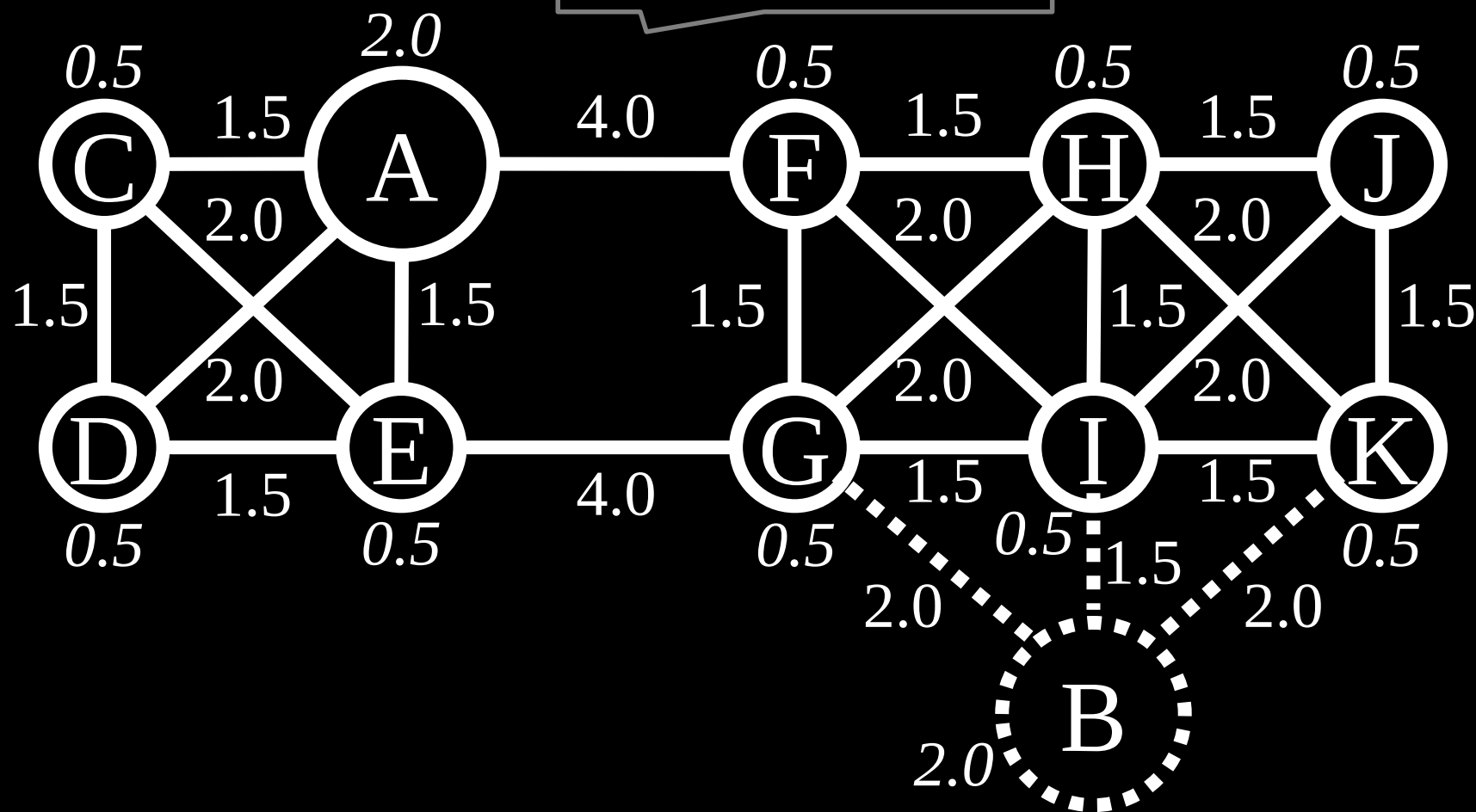
例3: 殖民型のトポロジー

東西間に隘路



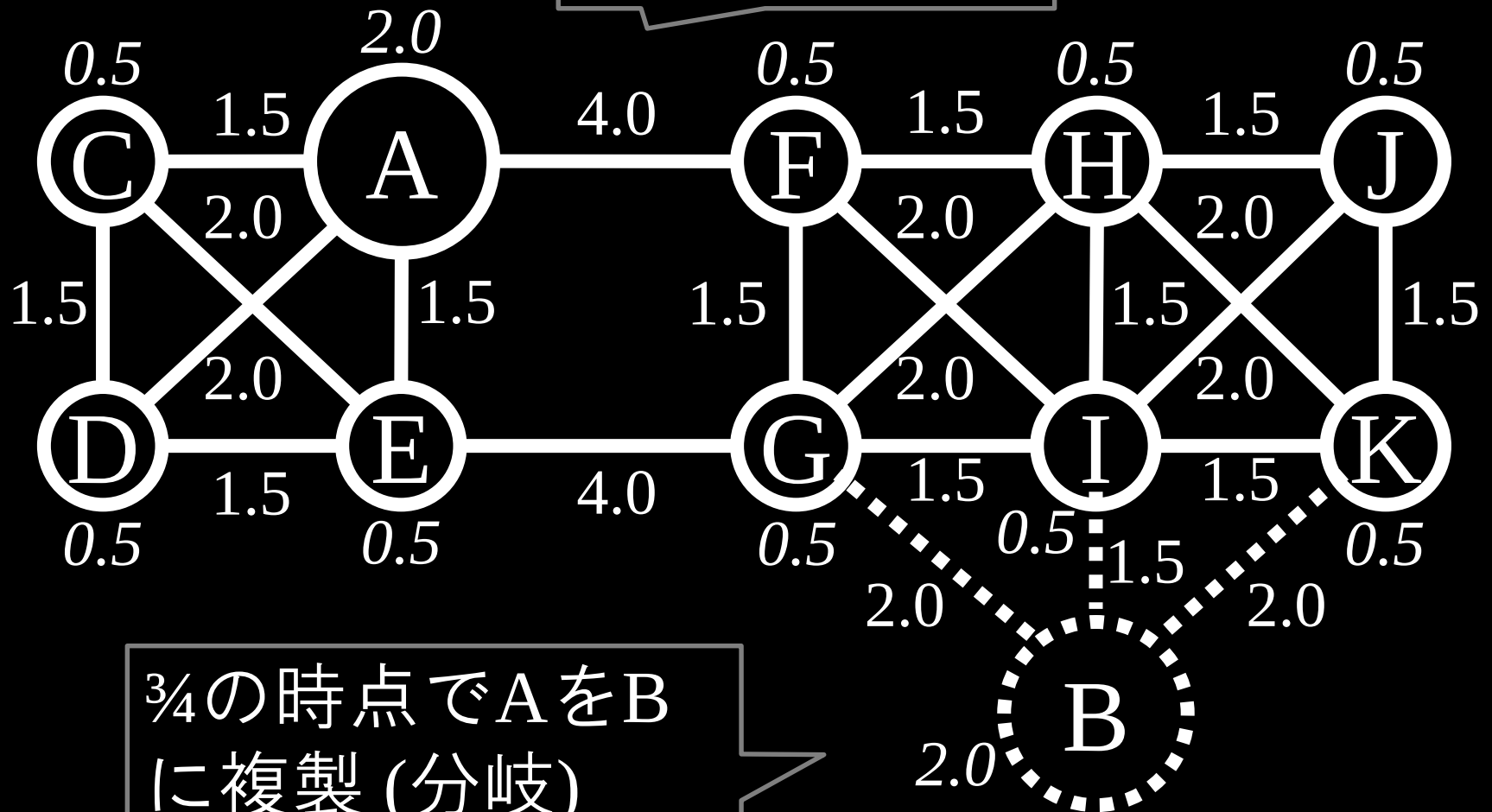
例3: 殖民型のトポロジー

東西間に隘路

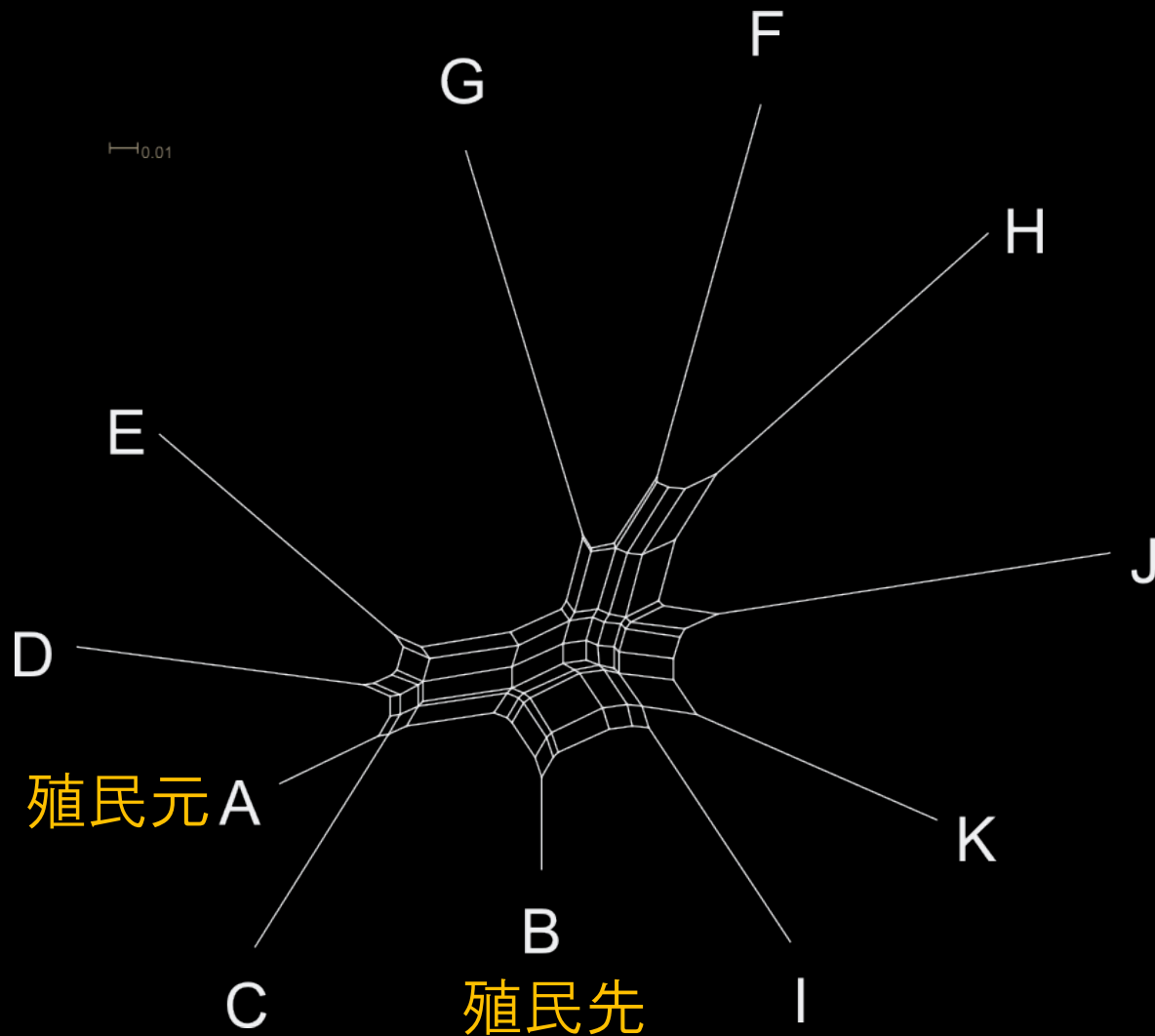


例3: 殖民型のトポロジー

東西間に隘路

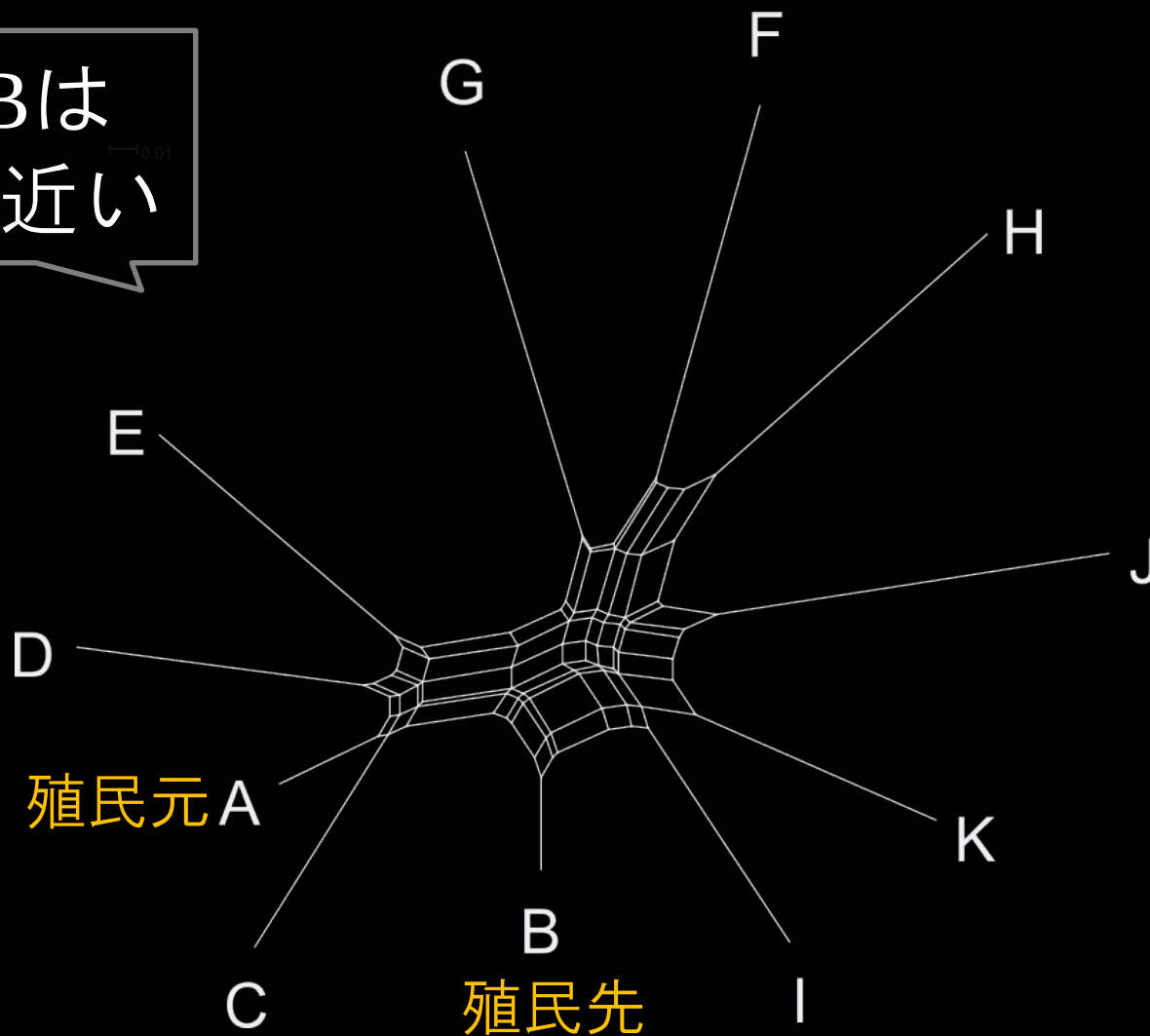


例4: 殖民型のNeighborNet



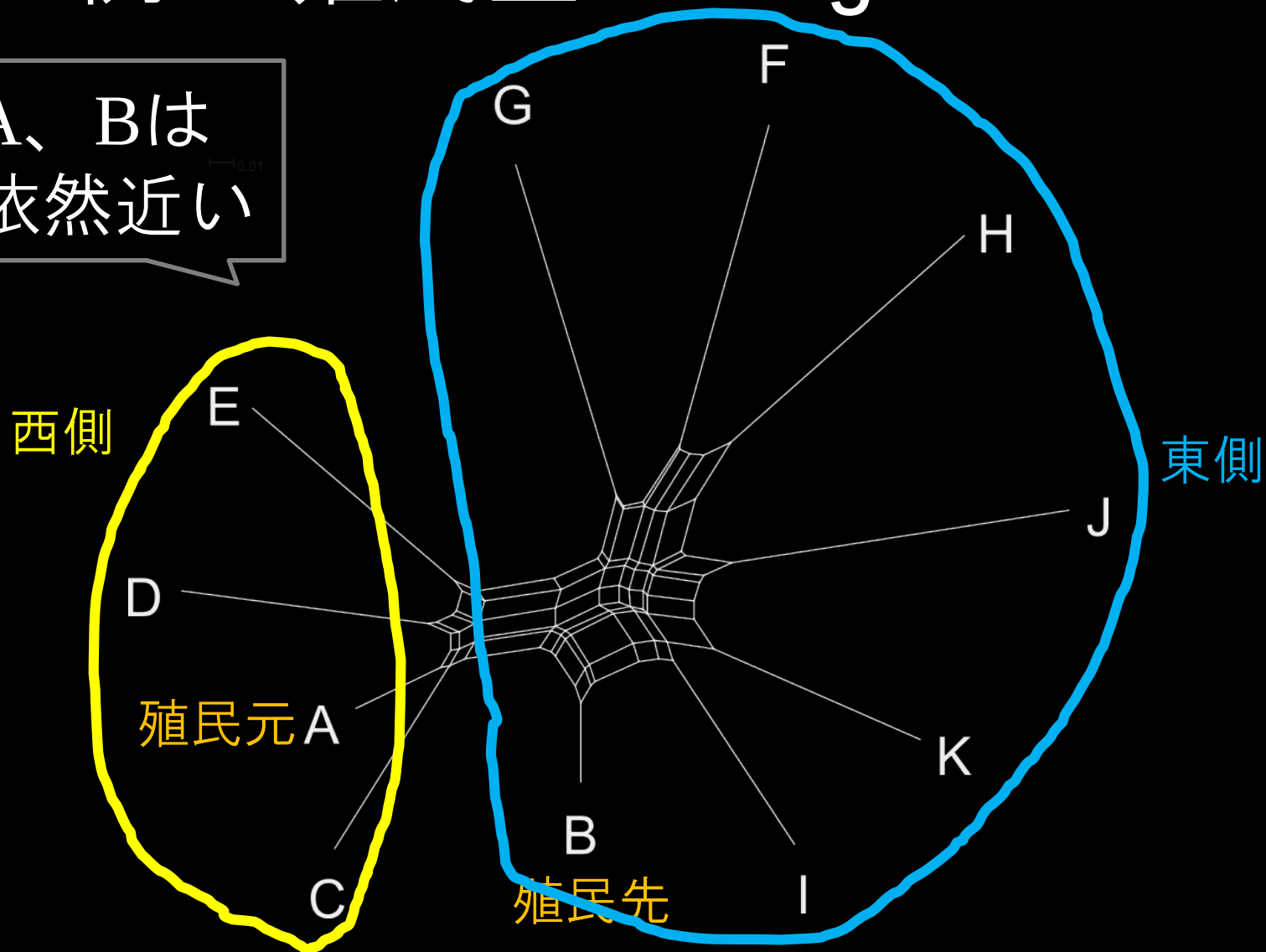
例4: 殖民型のNeighborNet

A、Bは
依然近い

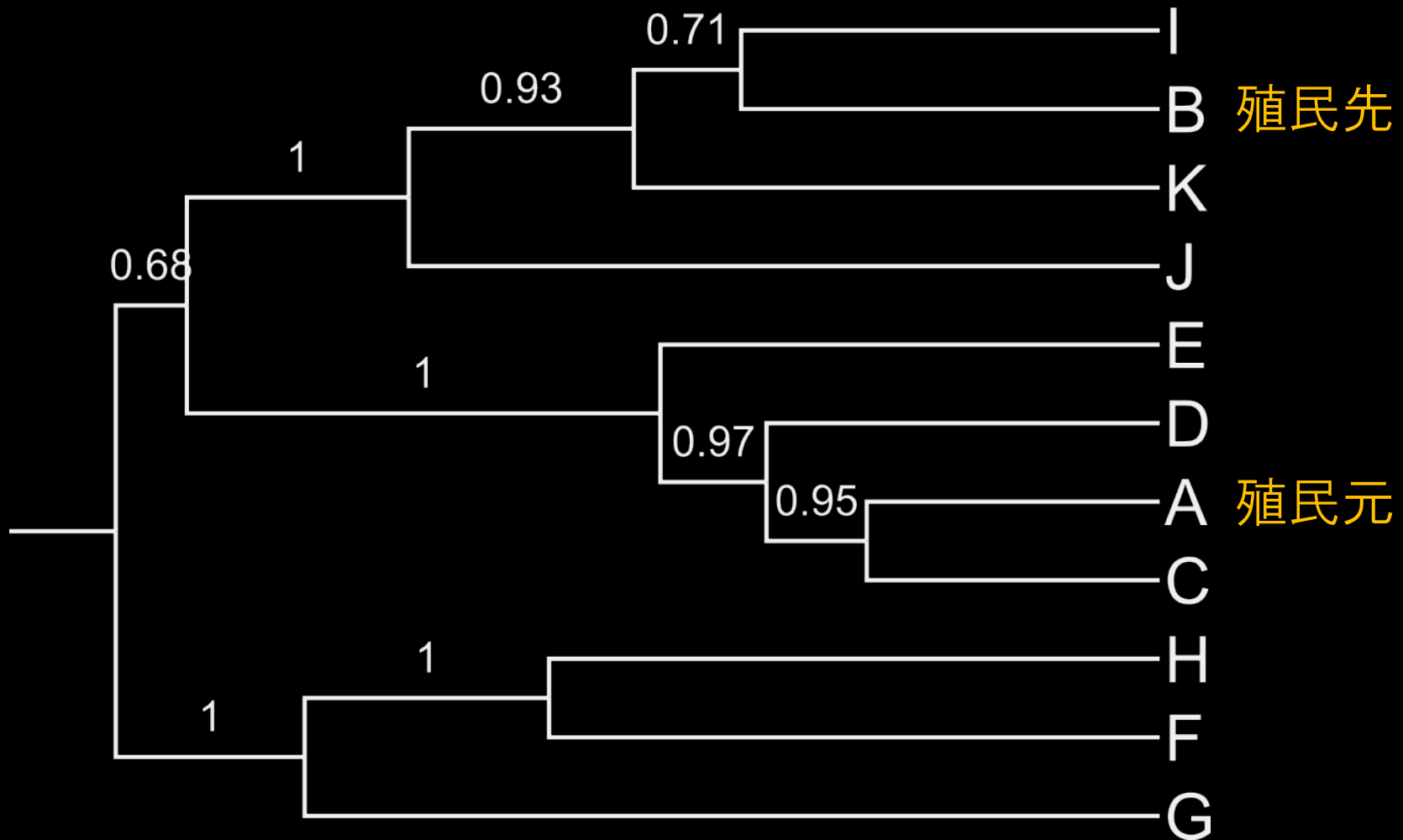


例4: 殖民型のNeighborNet

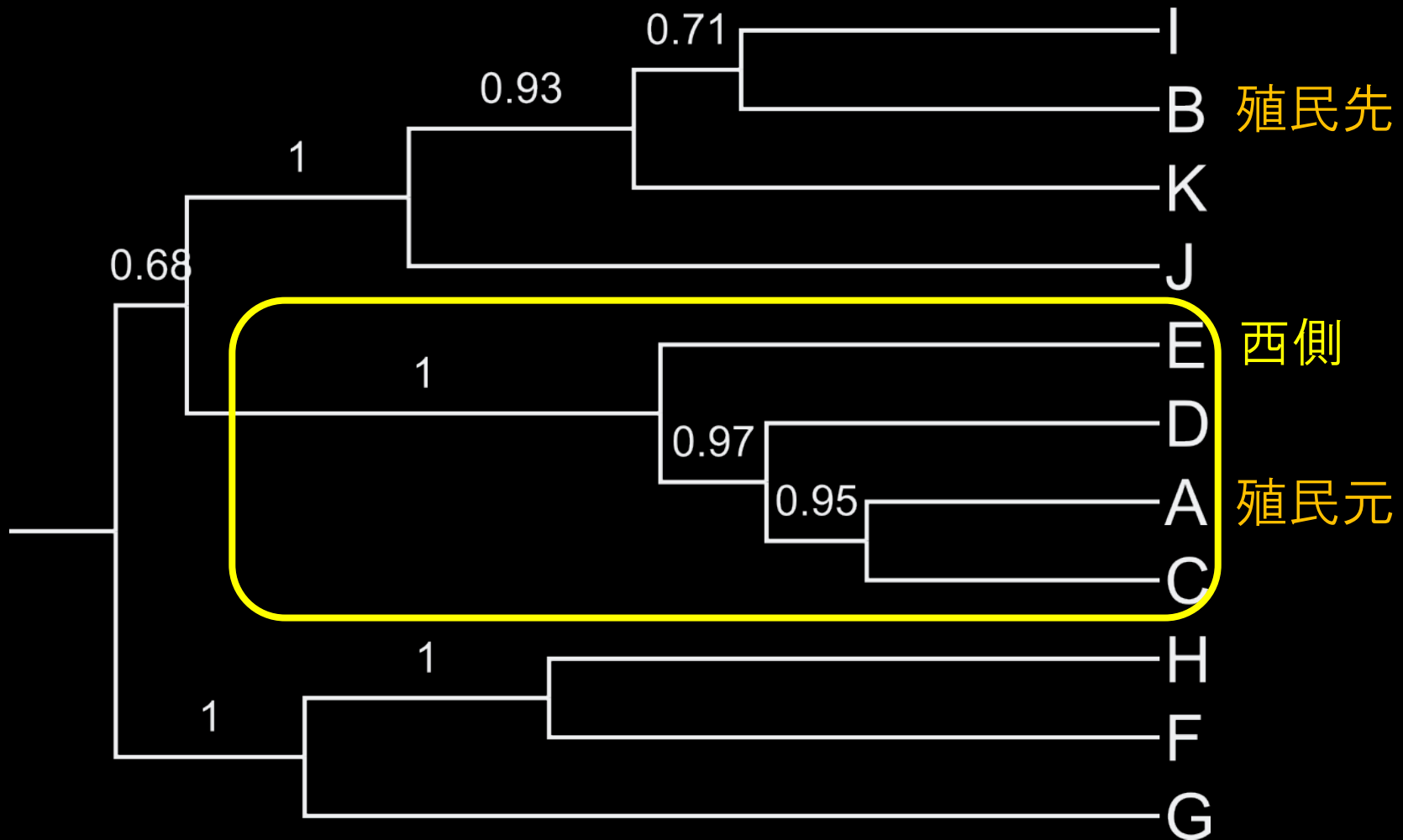
A、Bは
依然近い



例4: 殖民型の系統樹

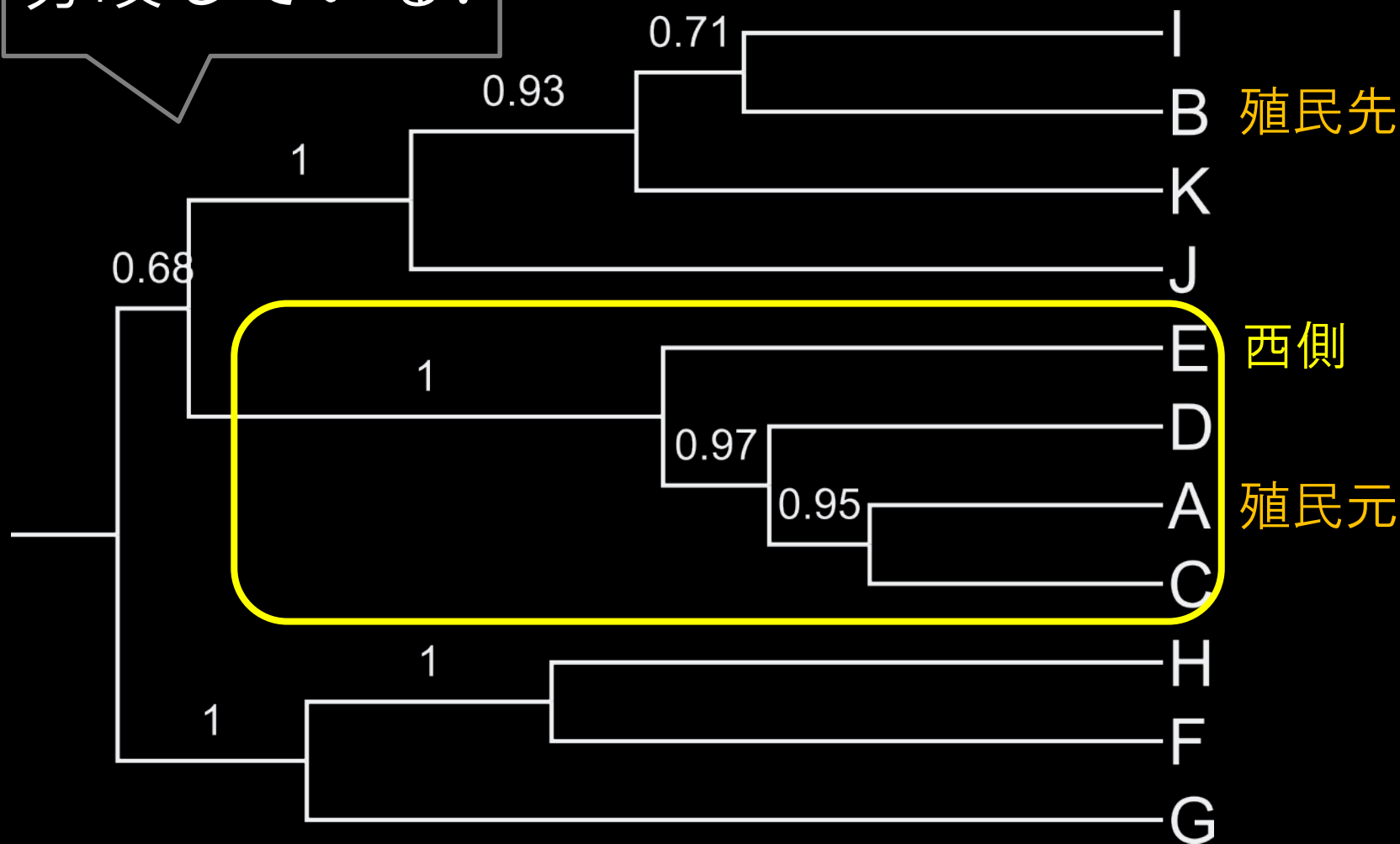


例4: 殖民型の系統樹



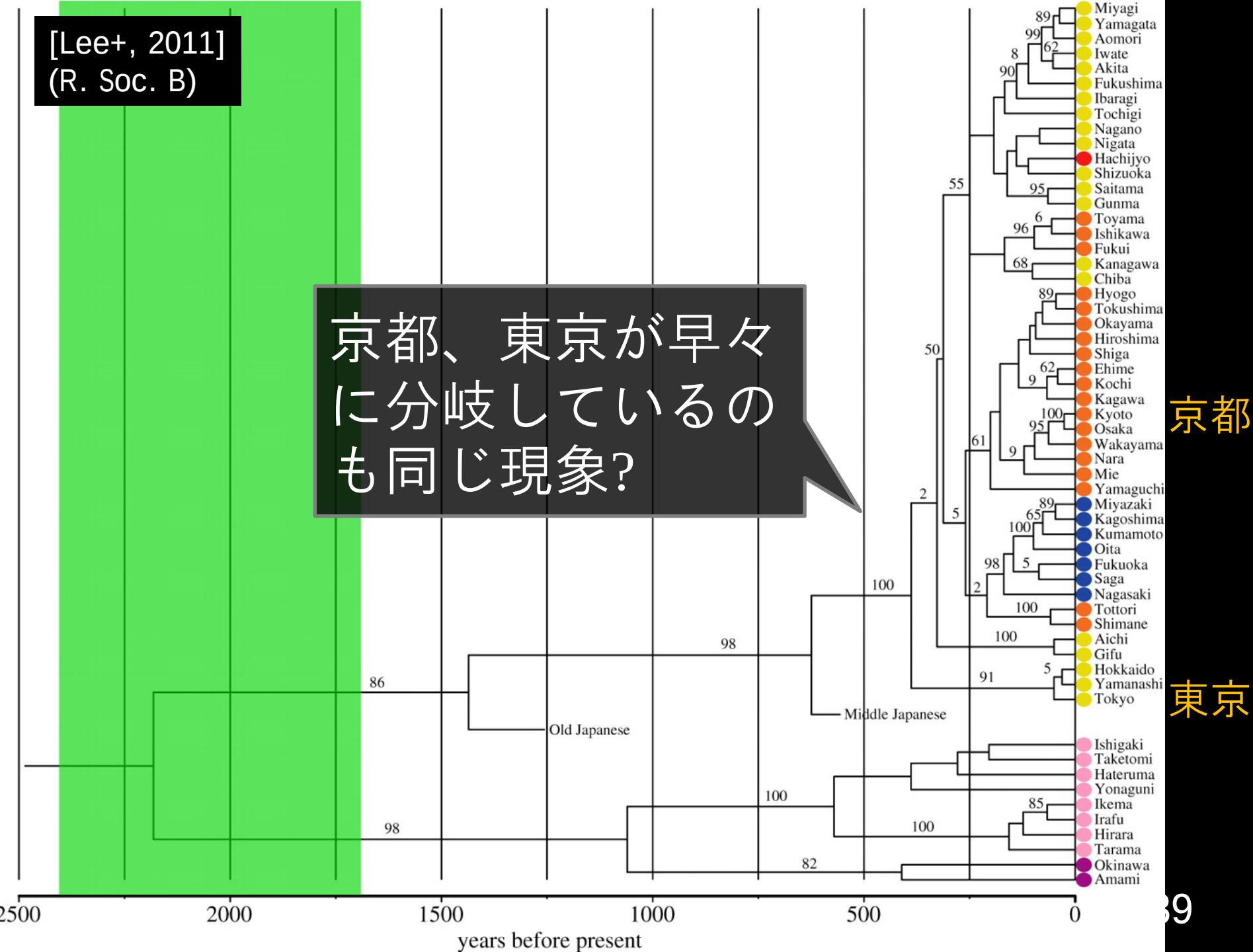
A、Bが早々に
分岐している!

殖民型の系統樹



[Lee+, 2011]
(R. Soc. B)

京都、東京が早々に分岐しているのも同じ現象?



京都

東京

まとめと今後の課題

- 接触に基づく語彙変化モデルを提案
- シミュレーションによる発見的、定性的な分析
- 今後の課題
 - パラメータが恣意的なので根拠が欲しい
 - 集団ではなく個体レベルでのモデル化
 - 言語の変化速度の違いを説明したい
 - どうしてアイスランド語は超保守的なのか
 - どうして琉球諸方言はあれほど変化が激しいのか

