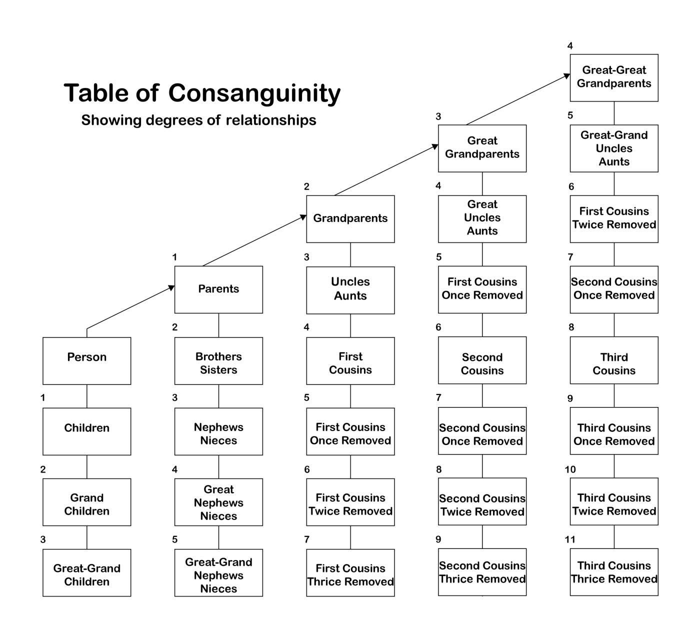
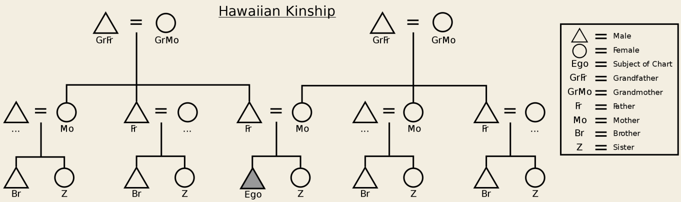
The Shortest Kinship Description Problem

Chao Xu, Qian Zhang
University of Illinois, Urbana-Champaign



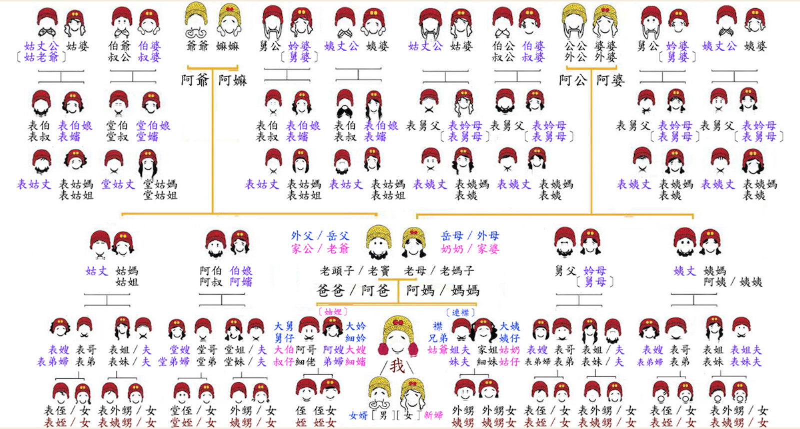
credit: wikipedia

Hawaiian Kin Terms



credit: wikipedia

Chinese kin terms



credit: wikipedia

A specially difficult situation...

- Relatives visit each other during Chinese New Year.
- Need to find the right word to address the other person.
- there are iphone apps for this.

A specially difficult situation...

- Relatives visit each other during Chinese New Year.
- Need to find the right word to address the other person.
- there are iphone apps for this.

Shortest Kinship Description Problem

Given the relation between ego and a kin, find a short description.

The kinship monoid

$\Sigma = \{f, m, d, s\}$.

i.e. *ffm* denotes father's father's mother.

Generation function:

$\mathrm{sgn}(f) = \mathrm{sgn}(m) = +$, $\mathrm{sgn}(d) = \mathrm{sgn}(s) = -$.

Gender function:

$g(m) = g(d) \neq g(f) = g(s)$.

Relators

The relators Γ consists of the following:

- generational cancellation:
  $abc = c$ if $\mathrm{sgn}(abc) \in \{+-+, -+-\}$,
- child's child's parent is child if gender agrees:
  $abc = a$ if $\mathrm{sgn}(abc) = --+$ and $g(a) = g(c)$,
- sibling/spouse:
  $ac = bc$ if $\mathrm{sgn}(ac) = \mathrm{sgn}(bc)$ and $\mathrm{sgn}(a) \neq \mathrm{sgn}(c)$.

$K = \langle \Sigma \mid \Gamma \rangle$ is the kinship monoid.

Problem (Submonoid membership optimization problem (SMOP))

For a monoid M, a subset $S \subseteq M$ and $x \in M$. Find a shortest sequence $x_1, \ldots, x_n \in S$, such that $\prod_i x_i = x$.

The shortest kinship description problem(SKDP) is SMOP over K.

$K = \langle \Sigma \mid \Gamma \rangle$ is the kinship monoid.

Problem (Submonoid membership optimization problem (SMOP))

For a monoid M, a subset $S \subseteq M$ and $x \in M$. Find a shortest sequence $x_1, \ldots, x_n \in S$, such that $\prod_i x_i = x$.

The shortest kinship description problem(SKDP) is SMOP over K. S is the set of *atomic kin terms* in the language.

Computational prespective

Input is given as strings over Σ. $X^* = \{ww' | w \in X^*, w \in X\}$. S is given as $W = \{w_1, \ldots, w_n\} \subseteq \Sigma^*$, x is given as $w \in \Sigma^*$. W the set of *atomic terms*, and W^* are the *terms*. Let $[w]$ be all strings in Σ^* that is equivalent to w. Let the cost of an element $w \in W^*$ to be the shortest sequence of strings in W that concatenates to w.

$$\$(w) = \min\{n | w_1 \ldots w_n = w, w_i \in W\}$$

To solve for SKDP, we are looking for a string in $[w] \cap W^*$ with minimum cost.

As a shortest path problem in a graph

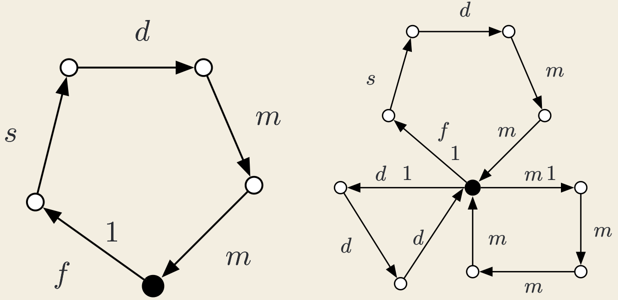


Figure: On the left is C_a, where $a = \texttt{fsdmm}$ and cost of a is 1. On the right is $G_{W,\$}$ where $W = \{\texttt{fsdmm}, \texttt{ddd}, \texttt{mmmm}\}$, and the cost of each word is 1. The black vertex is the special start vertex q. Edges without cost label have cost 0. All qq-paths has label in W^*.

Rewriting System

A rewriting system R is a subset of $\Sigma^* \times \Sigma^*$. If $(a, b) \in R$, we write $a \longrightarrow b$ to emphasis it is a rule.
Let $a \longrightarrow b \in R$. For $w = uav$ and $w' = ubv$, then we write $w \longrightarrow_R w'$.
We write $w \longrightarrow_R^* w'$ if there is a sequence of application rules in R to obtain w' from w.
Note $[w] = \left\{ w' | w \longrightarrow_\Gamma^* w' \right\}$.

For a rewriting system R with $L \subseteq \Sigma^*$.

- w is normal with respect to R if $w \rightarrow_R^* w'$ implies $w = w'$.
- w is the normal form of w' if $w' \rightarrow_R^* w$ and w is normal. $\underline{w}$ to denote the normal form of w.
- R is convergent if every string has a unique normal form.

The descandents of L: $\Delta_R^*(L) = \left\{w' | w \in L, w \xrightarrow{*}_R w'\right\}$.

The ancestors of L: $\nabla_R^*(L) = \left\{w | w' \in L, w \xrightarrow{*}_R w'\right\}$.

A convergent rewriting rule

- R the length reducing rules in Γ,
- $$S = \{fd \longrightarrow md, fs \longrightarrow ms, sf \longrightarrow df, sm \longrightarrow dm\} .$$

Let $T = S \cup R$. T is convergent. The symmetric closure of T is Γ, therefore:

Theorem

$[w] = \nabla_T^*(\underline{w})$

A convergent rewriting rule

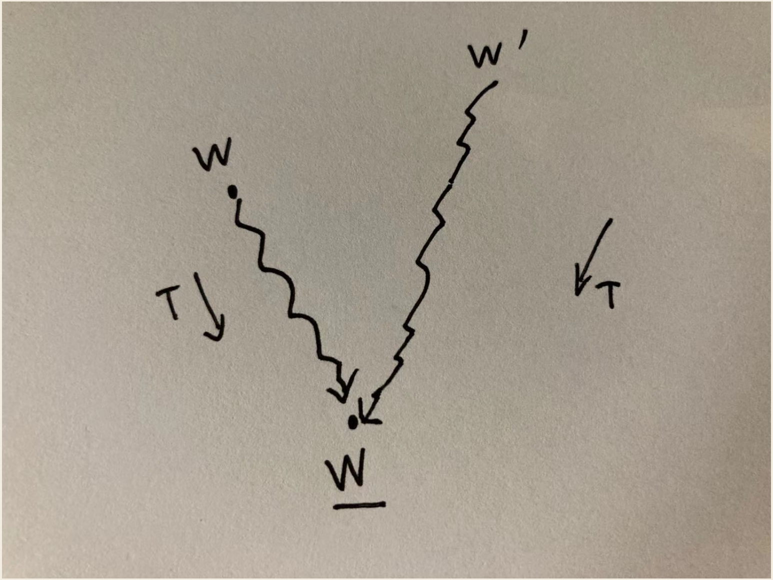
- R the length reducing rules in Γ,
- $$S = \{fd \to md, fs \to ms, sf \to df, sm \to dm\}.$$

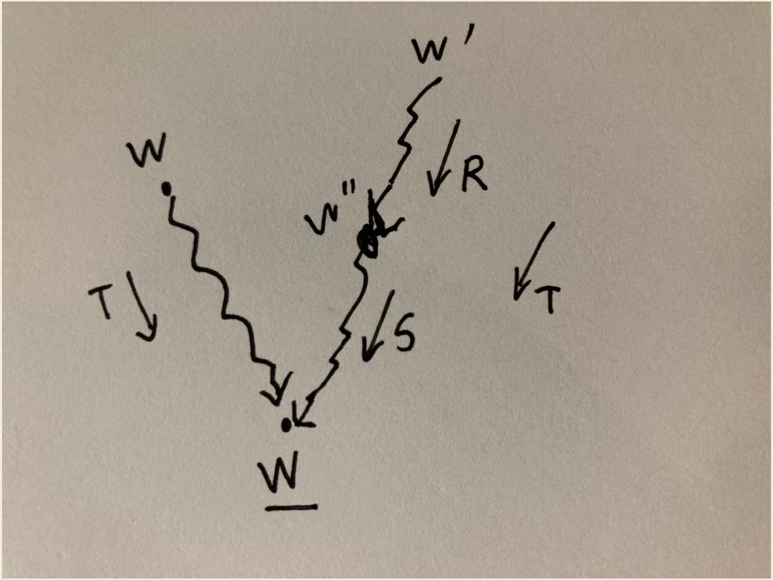
Let $T = S \cup R$. T is convergent. The symmetric closure of T is Γ, therefore:

Theorem

$[w] = \nabla_T^*(\underline{w})$

Original problem: Find a minimum cost string in $\nabla_T^*(\underline{w}) \cap W^*$.


w
w'
T
T
W


w'
w
R
w''
T
T
S
w

Theorem

For any $w \in \Sigma^*$, $\nabla_T^*(\underline{w}) = \nabla_R^*(\nabla_S^*(\underline{w}))$.

Theorem

For any $w \in \Sigma^*$, $\nabla^*_T(\underline{w}) = \nabla^*_R(\nabla^*_S(\underline{w}))$.

Find a minimum cost word in:

1. $[w] \cap W^*$,

Theorem

For any $w \in \Sigma^*$, $\nabla_T^*(\underline{w}) = \nabla_R^*(\nabla_S^*(\underline{w}))$.

Find a minimum cost word in:

1. $[w] \cap W^*$,
2. $\nabla_T^*(\underline{w}) \cap W^*$,

Theorem

For any $w \in \Sigma^*$, $\nabla_T^*(\underline{w}) = \nabla_R^*(\nabla_S^*(\underline{w}))$.

Find a minimum cost word in:

1. $[w] \cap W^*$,
2. $\nabla_T^*(\underline{w}) \cap W^*$,
3. $\nabla_S^*(\underline{w}) \cap \Delta_R^*(W^*)$,

New cost:

$$\$_R(w) = \min_{\substack{w' \xrightarrow{*}_R w \\ w' \in W^*}} \$(w')$$

How to solve it

How to solve it

- A graph G', labels of qq-paths is $\Delta_R^*(W^*)$, and the cost of a qq-path with label w is cost equals $\$_R(w)$.

How to solve it

- A graph G', labels of qq-paths is $\Delta_R^*(W^*)$, and the cost of a qq-path with label w is cost equals $\$_R(w)$.
- G' can be computed using a general construction [Myasnikov, Nikolaev, Ushakov 2015]. For our special case, in polynomial time due to context-free grammar.

How to solve it

- A graph G', labels of qq-paths is $\Delta_R^*(W^*)$, and the cost of a qq-path with label w is cost equals $\$_R(w)$.
- G' can be computed using a general construction [Myasnikov, Nikolaev, Ushakov 2015]. For our special case, in polynomial time due to context-free grammar.
- Find the shortest qq-path in G' with label in $\nabla_S^*(\underline{w})$.

How to solve it

- A graph G', labels of qq-paths is $\Delta_R^*(W^*)$, and the cost of a qq-path with label w is cost equals $\$_R(w)$.
- G' can be computed using a general construction [Myasnikov, Nikolaev, Ushakov 2015]. For our special case, in polynomial time due to context-free grammar.
- Find the shortest qq-path in G' with label in $\nabla_S^*(\underline{w})$.
- Finding a shortest path with labels in a regular language has a well known polynomial time algorithm. [Barrett, Jacob, Marathe 2000].

Theorem (Main)

For a input of total length n, the shortest kinship description problem can be solved in $O(n^3 + s)$ time, where s is the length of the shortest sequence.

Conjecture

There exists a polynomial p, such that for any W with total length n, and generator $x \in \{f, m, d, s\}$, either $[x] \cap W^*$ is empty, or there exists a sequence of $O(p(n))$ elements in W, such that its product is in $[x]$.