Wykład IX: Drzewa konsensusu i superdrzewa

Anna Gambin
28.11-5.12.2002

1 Definicja drzewa konsensusu

Zastanowimy się nad następującm zagadnieniem: jak dla rodziny drzew zbudować jedno *drzewo konsensusu*, które będzie zawierać sumaryczną informację zawartą w całej rodzinie.

Istnieje wiele przyczyn, dla których dysponujemy zbiorem różnych drzew (a nie jednym). Dla ustalonego zbioru gatunków możemy rozważać wiele drzew zbudowanych na podstawie różnych genów.

Często różne metody rekonstrukcji drzewa zastosowane do tych samych danych dają nam nieidentyczne wyniki. Niektóre metody (np. metoda parsymonii) nie potrafią w rozsądnym czasie zdecydować się na jedno optymalne drzewo, tylko podają cały suboptymalny zbiór.

Istnieje wiele pomysłow jak zdefiniować drzewo konsensusu.Prezentujemy niektóre z nich za [1] (w wersji dla drzew ukorzenionych, większość ma też nieukorzenione odpowiedniki):

1.1 Dokładny konsensus

Jedna z najprostszych metod: dla rodziny drzew ukorzenionych $\mathcal{T}$ drzewo dokładnego konsensusu CT zawiera tylko te klastry, które są wspólne dla wszystkich drzew z rodziny:

$$\delta(CT) = \cap_{T\in\mathcal{T}}\delta(T)$$

1.2 Konsensus większościowy

Niech $\mathcal{T}$ będzie rodziną drzew. Rozważmy zbiór klastrów M, które są obecne w więcej niż połowie drzew. Zauważmy, że dla każdej pary klastrów istnieje co najmniej jedno drzewo, które zawiera je obydwa, czyli zbiór M jest kompatybilny (bo jest parami kompatybilny). Drzewo konsensusu większościowego zawiera dokładnie te klastry, które należa do M.

1.3 Konsensus Adamsa

To jedna ze starszych i bardziej popularnych metod. Podamy algorytm konstruujący drzewo Adamsa, a później rozważymy podstawowe własności tego drzewa. Rozważmy rodzinę podziałów $\pi_1, \pi_2, \ldots \pi_k$ zbioru (liści) L. **Podział produktowy** L dla tej rodziny podziałów umieszcza dwa elementy w tym samym bloku, jeśli są one w tym samym bloku we wszystkich podziałach $\pi_1, \pi_2, \ldots \pi_k$.

Drzewo Adamsa powstaje rekurencyjnie na podstawie podziałow produktowych w następujący sposób:

Algorithm 1 Konsensus Adamsa

1. Jeśli drzewo T_1 zawiera tylko jeden liść to zwróć to drzewo.
2. wpp, dla każdego $i = 1, 2, \ldots k$, niech π_i będzie podziałem zbioru liści indukowanym przez maksymalne poddrzewa T_i.
3. Niech π podział produktowy dla $\pi_1, \pi_2, \ldots \pi_k$.
4. Dla każdego bloku B podziału π skonstruuj Konsensus Adamsa dla $T_{1|B}, T_{2|B}, \ldots T_{k|B}$.
5. Podłącz otrzymane drzewa do wspólnego korzenia i zwróć to drzewo.

Można pokazać, że drzewo Adamsa zawiera całą informację o zagnieżdżeniach wspólnych dla zbioru drzew. Dodatkowo dla każdej pary klastrów X, Y w drzewie Adamsa, takich że $X \subset Y$ mamy również $X \subset Y$ w każdym z drzew T_i należących do naszej rodziny. Te dwie własności charakteryzują (jednoznacznie) drzewo konsensusu Adamsa.

Oprócz tego można pokazać, że konsensus Adamsa nie wprowadza nowej (nieobecnej w rozważanym zbiorze drzew) informacji w postaci ukorzenionych trójek.

Twierdzenie 1 *Niech CT będzie drzewem konsensusu Adamsa dla rodziny drzew $\mathcal{T} = \{T_1, T_2, \ldots T_k\}$. Wtedy*

$$r(CT) \subseteq \cup_i r(T_i).$$

Dowód: Weźmy $ab|c \in r(CT)$. Istnieje zbiór S, taki że podział produktowy indukowany przez maksymalne poddrzewa $T_{1|S}, T_{2|S}, \ldots T_{k|S}$ wrzuca a i b do jednego bloku a c do innego. Czyli znajdziemy drzewo T_i zawierające a i b w pewnym maksymalnym poddrzewie $T_{i|S}$ a c w innym. Stąd $ab|c \in r(T_i)$.

1.4 Konsensus Aho

Dla zbioru ukorzenionych drzew $\mathcal{T}$ możemy skonstruować zbiór wspólnych ukorzenionych trójek $R = \cap_{T \in \mathcal{T}} r(T)$ i zastosować algorytm Aho znajdujący drzewo należące do $< R >$. Udowodnimy wspomniany już związek drzewa Aho z drzewem Adamsa:

Twierdzenie 2 *Niech R będzie zbiorem ukorzenionych trójek nad zbiorem liści L. Drzewo Aho dla R jest tożsame z drzewem konsensusu Adamsa dla $< R >$.*

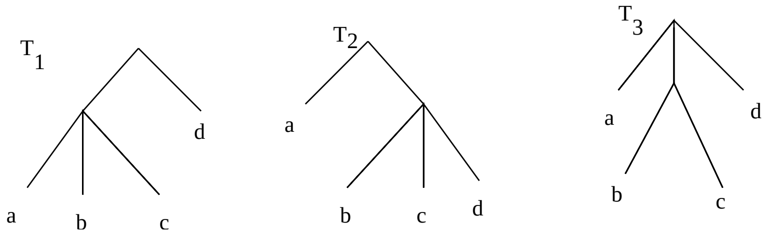
Dowód: Zarówno drzewo Aho jak i drzewo konsensusu Adamsa powstają rekurencyjnie — wystarczy więc udowodnić, że podział liści π_1 zdefiniowany przez maksymalne poddrzewa drzewa Aho dla R jest identyczny z podziałem π_2 zdefiniowanym przez maksymalne poddrzewa drzewa konsensusu Adamsa dla $< R >$. Zauważmy, że drzewo Aho $AT \in < R >$ czyli podział π_2 jest podziałem produktowym podziału π_1 i innych podziałów odpowiadających drzewom z $< R >$. Czyli π_2 jest uszczegółownieniem π_1.

Załóżmy teraz, że a i b należą do tego samego bloku π_1, czyli leżą w tej samej spójnej składowej grafu $[R, \mathcal{L}(R)]$, a więc są w tym samym maksymalnym poddrzewie dla dowolnego drzewa z $< R >$. Jeżeli a i b należą do tego samego maksymalnego poddrzewa we wszystkich drzewach z $< R >$ to należą do tego samego bloku podziału π_2, czyli $\pi_1 = \pi_2$.

Powyższe twierdzenie nie zachodzi dla dowolnego zbioru drzew ukorzenionych R (por. Rysunek 1)

1.5 Konsensus Nelsona

Oryginalna definicja drzewa konsensusu Nelsona jest niejednoznaczna. Podamy zmodyfikowaną wersję zaproponowaną przez Page'a. Rozważmy multizbiór klastrów wszystkich drzew z rodziny $\mathcal{T}$. Każdemu klastrowi nada-



Rysunek 1: Drzewo konsensusu Adamsa dla T_1 oraz T_2 jest równe T_3, natomiast drzewo Aho dla nich jest gwiazdą (nie mają wspólnych trójek, ale mają wspólne zagnieżdżenia)

jemy teraz wagę proporcjonalną do krotności jego wystąpień w tym zbiorze i tworzymy zwykły zbiór klastrów K. Drzewo konsensusu Nelsona jest zdefiniowane poprzez maksymalny wagowo kompatybilny podzbiór K. Kiedy występuje więcej niż jeden kompatybilny podzbiór o maksymalnej wadze drzewo jest zdefiniowane przez przecięcie maksymalnych zbiorów klastrów.

1.6 Konsensus mediany

Pomysł polega na znalezieniu "środkowego" drzewa dla zbioru drzew, czyli drzewa T które minimalizuje sumaryczną odległość od całej rodziny $\mathcal{T}$:

$$d_s(T, \mathcal{T}) = \sum_i d_s(T, T_i).$$

Najczęściej rozważaną odległością między dwoma drzewami jest w tym przypadku miara różnicy symetrycznej, czyli:

$$d_s(T_1, T_2) = |\delta(T_1) \setminus \delta(T_2)| + |\delta(T_2) \setminus \delta(T_1)|.$$

Jeżeli rozważymy nieparzystą liczbę drzew, wtedy tak zdefiniowany konsensus mediany jest równy drzewu konsensusu większościowego. Dla parzystej liczby drzew konsesnsus większościowy jest nadal tożsamy z konsensusem mediany, ale drzewami konsensusu mediany są również drzewa zdefiniowane przez zbiór klastrów, który powiększymy o te występujące dokładnie w połowie drzew z rodziny $\mathcal{T}$.

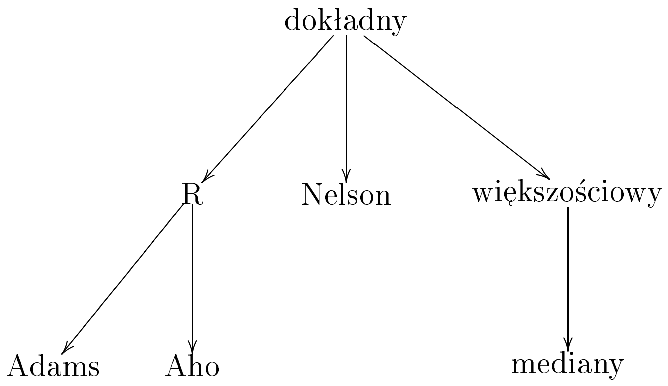
Rozważa się też asymetryczny konsensus mediany T, który minimalizuje:

$$\sum_i |\delta(T_i) \setminus \delta(T)|.$$

Dla dwóch drzew konsensus mediany można znaleźć w czasie wielomianowym (jak ?). Ogólny problem konsensusu mediany dla rodziny drzew (nawet binarnych) jest NP trudny.

1.7 Klasyfikacja drzew konsensusu, podejście aksjomatyczne

Postaramy sie podać klasyfikację wymienionych metod konsensusu. Zazwyczaj porównuje się ilość informacji, którą zachowuje drzewo konsensusu z informacją zachowywaną przez dokładny konsensus. W naszej klasyfikacji użyjemy porządku $\trianglelefteq$; przypomnijmy że $T_1 \trianglelefteq T_2$ (czyli T_2 rozszerza T_1) jeśli T_1 jest indukowanym poddrzewem pewnej kontrakcji T_2. W poniższym diagramie strzałka od pewnej metody konsensusu C_1 do innej (C_2) oznacza, że dla dowolnej rodziny drzew drzewo konsensusu stworzone metodą C_2 zawsze rozszerza drzewo konsensusu (lub każde z drzew konsensusu) zwracane przez metodę C_1. Węzeł R nie jest metodą konsensusu — oznacza zbiór wspólnych trójek ukorzenionych.



Można pokazać, że niekiedy metody zwracające różne drzewa dla dowolnej rodziny drzew redukują się do tej samej metody w przypadku tylko dwóch drzew. Np konsensus większościowy jest równoważny dokładnemu i konsensusowi Nelsona.

Do zagadnienia konsensusu można podejść z dwóch stron. Zapreznatowane podejście polega na zaproponowaniu metody a następnie badaniu jej własności matematycznych.

Podejście dualne to zdefiniowanie listy pożądanych własności i próba konstrukcji metody, która je spełnia.

Załóżmy, że c jest funkcją konsensusu, która dla zbioru drzew, zwraca ich drzewo konsensusu. Niech $P = \{T_1, T_2, \ldots T_k\}$, $P' = \{T'_1, T'_2, \ldots T'_k\}$, $L = \mathcal{L}(T_i)$ oznaczmy przez $T-X$ drzewo o klastrach $\delta(T)-\{X\}$, oraz odpowiednio $P - X = \{T_1 - X, T_2 - X, \ldots T_k - X\}$.

Wymienimy kilka aksjomatów:

A_1 $c(\{T\}) = T$.

A_2 Jeśli $P_{|X} = P'_{|X}$ to $c(P)_{|X} = c(P')_{|X}$.

A_3 $\forall\ X \subseteq \delta(T)\ c(P_{|X}) = c(P)_{|X}$.

A_4 Jeżeli $X \in \cap_i \delta(T_i)$ to $X \in \delta(c(P))$.

A_5 $\delta(c(P)) \subseteq \cup_i \delta(T_i)$.

A_6 Jeżeli $P_{|X} - X = P'_{|X} - X$ to $c(P)_{|X} - X = c(P')_{|X} - X$.

A_7 $\ldots$

Można pokazać wiele logicznych zależności pomiędzy aksjomatami. Niektóre z nich są sprzeczne, np. nie istnieje c spełniające jednocześnie A_4 oraz A_6. Reasumując podejście konstruktywne (najpierw zdefiniuj metodę potem zastanów się jakie ma ona własności) jest może mniej eleganckie dla matematyka, ale za to o wiele bardziej użyteczne dla biologa, bo zaowocowało mniej lub bardziej sensownymi, ale za to wykorzystywanymi w praktyce metodami.

1.8 Konsensus maksymalnej kliki

Zastanowimy sie nad ujednoliconym podejściem definiującym konsensus. Rozważmy następującą metodę: dla rodziny drzew ważonych $\mathcal{T} = \{T_1, T_2, \ldots T_k\}$, skonstruuj zbiór S zawierający wszystkie klastry z $\mathcal{T}$. Każdemu klastrowi przypisz wagę będącą sumą odpowiednich krawędzi w drzewach z rodziny $\mathcal{T}$.

Dwa klastry w zbiorze S połącz krawędzią wttw jak są one kompatybilne. Znajdź maksymalną klikę w takim grafie. Odpowiada ona **konsensusowi maksymalnej kliki**, o którym można pokazać, że zawiera w sobie kilka poprzednio zdefiniowanych metod. Niech X będzie klastrem, oznaczmy przez $f(X, \mathcal{T}) = |\{T \in \mathcal{T} : X \in \delta(T)\}|$. W zależności od wagi jaką przypiszemy klastrowi X dostaniemy metodę:

- **dokładnego konsensusu**, jeśli każdy klaster ważony jest jako $f(X, \mathcal{T}) - k + \frac{1}{2}$.
- **większościowego konsensusu**, jeśli waga przypisana klastrowi wynosi $f(X, \mathcal{T}) - \frac{k}{2}$.
- **asymetrycznego konsensusu mediany**, dla wag wynoszących $f(X, \mathcal{T})$.

Przypomnijmy, że znalezienie maksymalnego kompatybilnego podzbioru jest problemem NP trudnym. Zauważmy jednak, że zagadnienie konsensusu maksymalnej kliki jest bardziej specyficzne, bo wiemy że rozważane w zbiorze S klastry pochodzą z rodziny drzew $\mathcal{T}$. Możnaby mieć nadzieje, że ograniczony rozmiar rodziny $\mathcal{T}$ pozwoli na uzyskanie wielomianowych algorytmów. Niestety okazuje się to prawdą tylko dla dwóch drzew w rodzinie $\mathcal{T}$. Sformułujmy formalnie problem **konsensusu maksymalnej kliki**:

> DANE: Rodzina drzew $\mathcal{T} = \{T_1, \ldots T_k\}$, funkcja wag $w : C \to \mathcal{R}$ gdzie $C = \cup_i \delta(T_i)$, liczba K.
>
> PYTANIE: Czy istnieje drzewo T, takie że $\delta(T) \subseteq C$ oraz
>
> $$\sum_{X \in \delta(T)} w(X) \geq K?$$

Twierdzenie 3 *Konsensus maksymalnej kliki jest wielomianowo rozstrzygalny dla $k \leq 2$ oraz NP trudny dla $k \geq 3$.*

2 Zagadnienie superdrzewa

Metody konsensusu rozważają rodzinę drzew, w której wszystkie drzewa mają ten sam zbiór liści. Superdrzewo konstruujemy dla rodziny $\mathcal{T}$, w której zbiory liści poszczególnych drzew nie muszą być identyczne [2].

Podejście superdrzewa wiąże się z dwoma podstawowymi problemami: (*i*) określić sensowne kryteria "łączenia" drzew; (*ii*) zaprojektować wielomianowy algorytm wykonujący to zadanie.

Pokazano w pracy [3], że nie istnieje sensowna metoda superdrzewa dla drzew nieukorzenionych. Z tego powodu zakłada się, że dysponujemy rodziną drzew ukorzenionych; przy czym korzeniem drzewa może być zarówno hipotetyczny wspólny przodek jak też liść wspólny dla wszystkich drzew w rodzinie.

Metody konstruujące superdrzewa możemy podzielić na dwie klasy: metody pośrednie i metody bezpośrednie. Metody bezpośrednie przypominają poznane przez nas metody konstruujące drzewo konsensusu. Konstruują one superdrzewo na podstawie, wejściowej rodziny drzew, zalicza się tutaj również tzw. metody nieformalne/ręczne.

Głowny pomysł metod pośrednich polega na zakodowaniu rodziny drzew $\mathcal{T}$ jako np. zbioru cech, a następnie budowaniu jednego "wspólnego" drzewa np. metodą parsymonii. Wszystkie rozważane tutaj podejścia są heurystyczne i nie istnieje sensowna miara dokładności tych metod. W dalszej części wykładu skoncentrujemy się na metodach bezpośrednich i dokładniej omówimy jeden algorytm budujący **superdrzewo minimalnego cięcia**.

2.1 Ograniczenia dla superdrzew

Okazuje się, że zarówno dla metod konstruujących superdrzewo jak i dla metod konstruujących drzewa konsensusu, istnieją pewne inherentne ograniczenia ich możliwości. Sformułujemy listę bardzo naturalnych własności jakie powinny speńiać te metody i pokażemy, że dla drzew nieukorzenionych nie istnieje metoda spełniająca je wszystkie.

Metody superdrzewa możemy zastosować do rodziny drzew, w której nie narzucamy żadnego porządku na elementach, jak również do uporządkowanego (liniowo lub częściowo) ciągu drzew. Porządek może oznaczać np. miarę sensowności/zaufania do danego drzewa. Jest jednak pożądane, żeby metoda mogła zostać zastosowana do dowolnego nieuporządkowanego zbioru drzew — to wymaganie tłumaczy własność (1) poniżej. Pozostałe dwie własności są też bardzo naturalne.

(1) metoda może zostać zastosowana do dowolnej rodziny drzew.

(2) Jeśli w drzewach wejściowych zmienimy bijektywnie nazwy gatunków, otrzymane drzewo jest identyczne z drzewem jakie skonstruuje metoda dla pierwotnego zbioru drzew, z dokładnością do przemianowania liści.

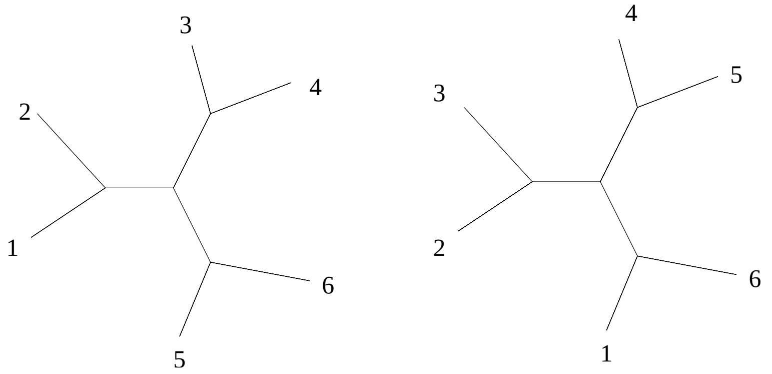
(3) Jeżeli rozważamy kompatybilną rodzinę drzew $\mathcal{T}$ (czyli $< \mathcal{T} > \neq \emptyset$) to drzewo zwrócone przez metodę powinno należeć do zbioru $< \mathcal{T} >$.

Zachodzi następujący fakt:

Fakt 1 *Dla zbioru drzew nieukorzenionych (nawet binarnych) nie istniej metoda konstruująca superdrzewo spełinająca własności* $(1) - (3)$.

Dowód: Rozważmy przykład trzech drzew dla 6 gatunków etykietowanych $1, 2, \ldots 6$. Niech $\mathcal{T} = \{T_1, T_2, T_3\}$ gdzie $T_1 = (12)(45)$, $T_2 = (34)(16)$ oraz $T_3 = (56)(23)$.

Zauważmy że te drzewa są kompatybilne. Załóżmy że metoda konstruująca superdrzewo spełnia własno:sci (1) i (2). Pokażemy, że nie może spełniać własności (3). Okazuje się, że zbiór $< \mathcal{T} >$ zawiera dokładniedwa drzewa (przedstawione na Rysunku 2).



Rysunek 2: Drzewa należące do $< \mathcal{T} >$.

Czyli metoda spełniająca własność (3), powinna zwrócić jako superdrzewo jedno z przedstawionych drzew. Załóżmy teraz, że zamieniamy nazwy gatunków następująco: 2 zamieniamy z 6 oraz 3 zamieniamy z 5. Po takiej zamianie nasz zbiór drzew pozostanie ten sam (zmieni się jedynie kolejność drzew). A więc nie powinno ulec zmianie drzewo zwracane przez metodę (o ile spełnia ona własności (1) i (2)). Niestety opisana zamiana nazw gatunków zmienia drzewa ze zbioru $< \mathcal{T} >$ (właściwie je zamienia). A więc nasza metoda nie może spełniać własności (3).

Wnioskujemy zatem, że trzeba albo zrezygnować z naszych własności, albo ograniczyć się do drzew ukorzenionych. Wybierzemy tę drugą możliwość i opiszemy algorytm bedący modyfikacją algorytmu Aho (który bada kompatybilność rodziny ukorzenionych drzew). Oprócz wspomnianych własności algorytm spełnia dwie dodatkowe:

(4) Każdy liść (gatunek) występujący co najmniej raz w rodzinie $\mathcal{T}$ wystąpi w superdrzewie.

(5) Czas konstrukcji superdrzewa jest wielomianowy względem liczby gatunków.

2.2 Ograniczenia dla metod konsensusu

Można pokazać, że większość poznanych przez nas metod konsensusu spełnia własności (1), (2), (4), (5) oraz, że może zostać zmodyfikowana, tak aby spełniać również (3).

Główny problem z drzewami konsensusu, polega na tym że dla rodziny, która nie jest kompatybilna drzewo konsensusu może zawierać bardzo mało informacji filogenetycznej. Pokażemy, że znów zadanie dla drzew nieukorzenionych jest tutaj trudniejsze, czyli nie istnieje dla nich odpowiednik konsensusu Adamsa (który zachowuje całą wspólną dla drzew informację). Chcielibyśmy żeby metoda spełniała następującą własność:

(6) Jeśli czwórka $ij|kl \in q(T_i)$ dla wszystkich drzew $T_i \in \mathcal{T}$ to ta czwórka zostanie zachowana w drzewie konsensusu.

Zachodzi następujący fakt:

Fakt 2 *Dla rodziny drzew nieukorzeniomych nie istnieje metoda konsensusu spełniająca własności* (1)*,* (2)*, oraz* (6)*.*

Dowód: Rozważmy znów dwa drzewa z Rysunku 2. Obydwa z nich zawierają czwórki $12|45, 34|16$, oraz $56|23$ i są jedynymi drzewami nad tym zbiorem czwórek. Czyli jeśli wymagamy własności (6) to drzewo konsensusu dla tych właśnie drzew musie być jednym z nich. Znowu zamieńmy gatunki 2 i 6 oraz 3 z 5. taka zamiana zamienia kolejność drzew ale jeśli metoda ma spełniać (1) oraz (2) nie powinna zmeieniać drzewa konsensusu. Niestety jak pamiętamy takie przemianowanie gatunków zamienia ze sobą nasze drzewa.

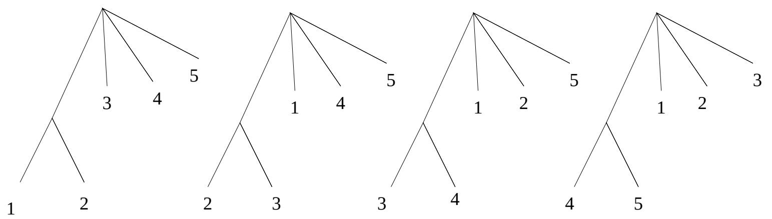
Zastanówmy się teraz nad drzewami ukorzenionymi. Sformułujmy dwie własności (pierwsza jest analogiczna do własności (6), druga jest dużo silniejsza):

(6') Jeśli wszystkie drzewa wejściowe zawierają trójkę $ij|k$ to drzewo konsensusu też ją zawiera.

(7) Jeśli chociaż jedno wejściowe drzewo zawiera trójkę $ij|k$ oraz żadne z wejściowych drzew nie zawiera trójek $ik|j$ ani $jk|i$ to drzewo konsensusu zawiera trójkę $ij|k$.

Fakt 3 *Dla rodziny drzew ukorzenionych istnieje metoda konsensusu spełniająca własności* $(1) - (5)$ *oraz* $(6')$*;*
natomiast nie istnieje metoda spełniająca własność (7).

Dowód: Metodą, która spełnia własności $(1) - (5)$ oraz (6) jest konsensus Adamsa. Aby udowodnić drugą część stwierdzenia rozważmy jako $\mathcal{T}$ trzy



Rysunek 3: Drzewa $T_1 - T_4$ ilustrujące niemożność spłnienia własności (7).

drzewa na Rysunku 3. Niech R będzie zbiorem trójek $\mathcal{T}$. Mamy $12|5 \in r(T_1)$, $23|5 \in r(T_2)$, $34|1 \in r(T_3)$ oraz $45|1 \in r(T_4)$. Widzimy, że żadna trójka postaci $ik|j$ lub $jk|i$ dla $ij|k \in \{12|5, 23|5, 34|1, 45|1\}$ nie należy do R. czyli drzewo konsensusu spełniające własność (7) musi zawierać trójki $TR = \{12|5, 23|5, 34|1, 45|1\}$. Jak łatwo sprawdzić nie istnieje drzewo ukorzenione T takie że $TR \subseteq r(T)$.

2.3 Superdrzewo minimalnego cięcia

Niech $\mathcal{T} = \{T_1, T_2, \ldots T_k\}$. Opiszemy dwa grafy wykorzystywane w algorytmie budującym superdrzewo dla $\mathcal{T}$.

Pierwszy z nich jest prawie identyczny z grafem budowanym w algorytmie Aho. Niech $S \subseteq \sum_{i=1}^{k} \mathcal{L}(T_i)$, oznaczmy przez $S_{\mathcal{T}}$ graf $(S, E_{\mathcal{T}})$ gdzie $(a, b) \in E_{\mathcal{T}}$ jeśli istnieje co najmniej jedno drzewo w $\mathcal{T}_{|S}$, dla którego a i b należą do wspólnego nietrywialnego klastra.

Drugi graf oznaczamy przez $S_{\mathcal{T}}/E_{\mathcal{T}}^{\max}$ i otrzymujemy z pierwszego w następujący sposób:

- Nadaj każdej krawędzi $(a, b) \in E_{\mathcal{T}}$ wagę równą liczności zbioru drzew, które zawierają obydwa liście w tym samym nietrywialnym klastrze.
- Dokonaj kontrakcji krawędzi, których waga równa się liczności zbioru rozważanych drzew (oznaczamy ten zbiór krawędzi przez $E_{\mathcal{T}}^{\max}$).
- Usuń wszystkie pętle a równoległe krawędzie zastąp krawędzią pojedynczą o wadze równej liczności zbioru tych drzew w rodzinie $\mathcal{T}_{|S}$, które posiadają nietrywialny klaster obejmujący wierzchołki incydentne do którejś z równoległych krawędzi.

Algorithm 2 Superdrzewo (T)

1. $S := \sum_{i=1}^{k} \mathcal{L}(T_i)$.
2. jeśli $|S| \leq 2$, zwróć drzewo o liściach z S.
3. wpp skonstruuj graf $S_{\mathcal{T}}$.
4. Niech $S_1, S_2, \ldots S_r$ będzie podziałem S indukowanym przez spójne składowe grafu $S_{\mathcal{T}}$
5. Jeżeli $r = 1$ czyli graf $S_{\mathcal{T}}$ jest spójny, skonstruuj graf $S_{\mathcal{T}}/E_{\mathcal{T}}^{\max}$. Niech E' będzie zbiorem krawędzi, które należą do co najmniej jednego minimalnego cięcia w grafie $S_{\mathcal{T}}/E_{\mathcal{T}}^{\max}$. Dla każdej krawędzi w zbiorze E' usuń odpowiednią krawędź (lub kilka krawędzi równoległych) z grafu $S_{\mathcal{T}}$. Otrzymasz w ten sposób podział S na składowe: $S_1, S_2, \ldots S_r$ gdzie $(r \geq 2)$.
6. Dla każdego $j \in \{1, 2, \ldots, r\}$ skonstruuj rekurencyjnie superdrzewo dla rodziny $\mathcal{T}_{|S_j} = \{T_{1|S_j}, T_{2|S_j}, \ldots T_{k|S_j}\}$.
7. Podłącz otrzymane drzewa pod wspólny korzeń.

Literatura

[1] D. Bryant. Hunting for trees, buiding trees and comparing trees: theory and method in phylogenetic analysis. *phd thesis, Department of Mathematics, University of Canterbury*, 1997.

[2] C. Semple and M. Steel. A supertree method for rooted trees. *Discrete Applied Mathematics*, 105:147–158, 2000.

[3] M. Steel, S. Böcker, and A.W.M. Dress. Some simple but fundamental limits for supertree and consensus tree methods. *Systematic Biology*, 42(2):363–368, 2000.

3 Zadania

1. Zaprojektuj wielomianowy algorytm znajdujący konsensus maksymalnej kliki dla dwóch drzew, oraz (również wielomianowy) algorytm znajdujący ten konsensus dla rodziny k ukorzenionych drzew binarnych.

2. Udowodni, że konsensus Adamsa spełnia własności (1) – (5) oraz (6′).