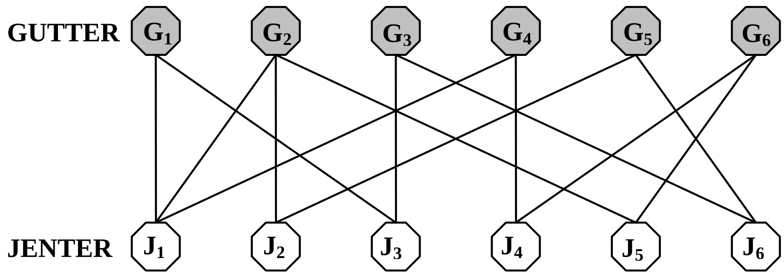
Online datingtjeneste – The Hungarian Algorithm

Problemet
Forestill deg at du har startet en online datingtjeneste hvor du lar brukerne sette opp en ønskeliste over hvilke andre brukere på siden de kunne tenke seg å gå ut med, de kan bare velge en bruker av motsatt kjønn. Du ønsker se på situasjonen som en graf, så du lager en graf hvor nodene er brukere og kantene er potensielle dater. I tillegg er nodene delt i to grupper som ikke har kanter innbyrdes, for å gjenspeile at man bare kan gå ut med en av motsatt kjønn.



Det har seg også slik at du er en sånn ”alt-eller-ingenting” type, så du ønsker at enten så får alle seg en date, eller så få ingen date. Med dette ekstra kravet er det åpenbart at det noen ganger ikke blir date på noen, f.eks hvis to gutter begge kun ønsker å gå ut med en jente. Problemet er nå å finne en slik ordning av dates mellom gutter og jenter, eller eventuelt bli sikker på at det ikke er mulig.

Koblinger
For å kunne løse problemet trenger vi å kunne snakke om at en gruppe brukere har fått seg en (om så midlertidig) datepartner, dette kaller vi en kobling (matching i boka). Så i eksempelet over så vil f.eks $\{(G_1,J_1), (G_2,J_5)\}$ være en kobling, mens $\{(G_1,G_1), (G_2,J_2), (G_5,J_2)\}$ åpenbart ikke er en kobling fordi G_1 ikke kan date seg selv, og fordi J_2 ikke kan gå ut med både G_2 og G_5. Løsningen på hele problemet er da å finne en kobling hvor alle brukerne inngår og da har fått seg en partner, som vi da kaller en perfekt kobling, eller finne en grunn til at det ikke går an å lage en kobling. Hva kan det tenkes at en slik grunn kan være?

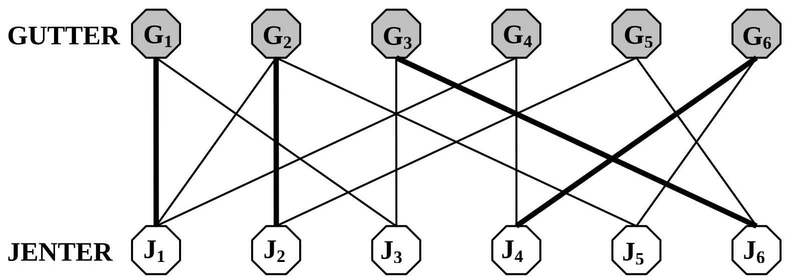
Den måten vi begrunner at det ikke går an å fikse date til alle er at vi peker ut en grunne gutter, og viser til at de bare kan tenke seg å gå ut med en gruppe jenter som er færre en dem. Det er åpenbart at hvis det blant alle dine brukere finnes 5 gutter som bare kunne tenke seg å date 3 jenter, så vil det aldri gå. Det er greit nok, men det som er litt forbausende er at hvis det ikke finnes noe slikt tilfelle, altså at for alle grupper gutter så kan de tenke seg å date en gruppe jenter som er minst like mange som dem, så finnes det garantert en kobling. Påstanden som sier dette er Hall’s Theorem.

Beviset for at Hall’s Teorem er sant er algoritmen The Hungarian Algorithm for Perfect Matching in Bipartite Graphs, og den vil nettopp peke på en gruppe gutter og si at disse er de som ødelegger for flertallet ved å ønske en mindre gruppe jenter enn seg selv. Navnet gjenspeiler akkurat det vi har tatt for oss, hvor matching betyr kobling, perfect gjenspeiler at vi vil ha en kobling hvor alle har en partner, bipartite betyr at grafens noder er delt i to grupper (gutter og jenter i vårt tilfelle) og selve grafen representerer potensielle dates mellom forskjellige brukerne.

Øke antall dates

Det man først ville ty til er å bare tilfeldig koble sammen gutter og jenter som vil ha hverandre, og deretter se om det fungere, men som forventet vil ikke dette føre frem. Er man heldig kan man klare å finne en datepartner for alle, men hvis det ikke fører frem så kan man ikke være sikker på om det ikke bare var noen uheldige valg som førte til at det ikke løste seg.

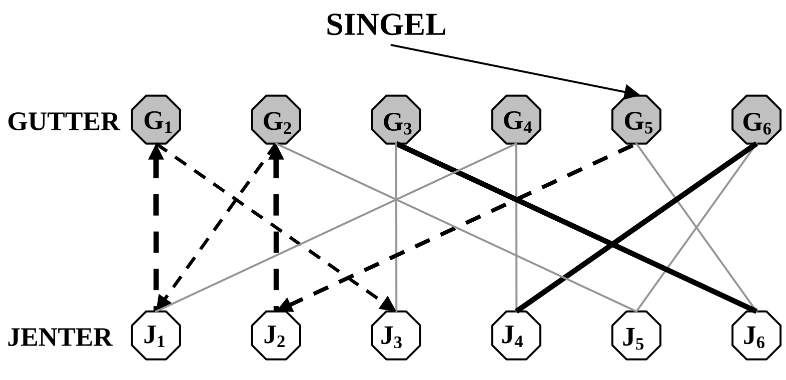
For å skjønne The Hungarian kan vi for enkelthets skyld anta at vi allerede har en kobling som ikke er perfekt, altså den inneholder ikke alle brukerne. Parene som er datepartnere er markert med tykk linje;



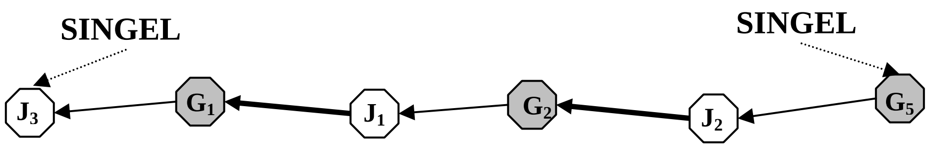
Vi ønsket å gjøre koblingen vår én date større, men som vi så tidligere kan vi ikke bare velge en tilfeldig mulig pardanning. Det vi trenger er en metode som klarer å gjøre koblingen én større, eller gir en grunn, som nevnt tidligere, for at en perfekt kobling er umulig. Som nevnte gjør The Hungarian akkurat dette, så la oss først se på hva som skjer når koblingen faktisk blir utvidet.

Forbedringsveier

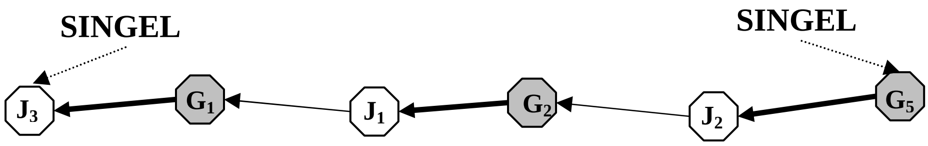
Man starter først med en gutt som er uten date, så går du til en av jentene han vil ha som allerede har en datepartner. Fra hun går til gutten som er hennes datepartner, og fra han til en jente som han vil ha, også videre. Annenhver gang går du mellom to som er datepartnere, og to som ikke er datepartnere men som valgte hverandre som aktuelle. Hvis du klarer å gå fra den ledige gutten i starten, og tilslutt ende opp hos en ledig jente, så kan du gjøre koblingen én date større ved å "invertere" parene. En slik vei gjennom grafen kan vi kalle en forbedringsvei gjennom koblingen, se eksempelet under:



Veien starter hos G_5 og slutter hos J_3, begge er single. Annenhver gang går vi langs et date/ikke-par. Først (G_5,J_2) som ikke skal på date, så (J_2,G_2) som skal på date, så (G_2,J_1) som ikke skal på date, osv. Her er et bilde av veien, legg merke til at første og siste bruker er begge single som nevnt;



Så inverterer vi datene og får en date ekstra;



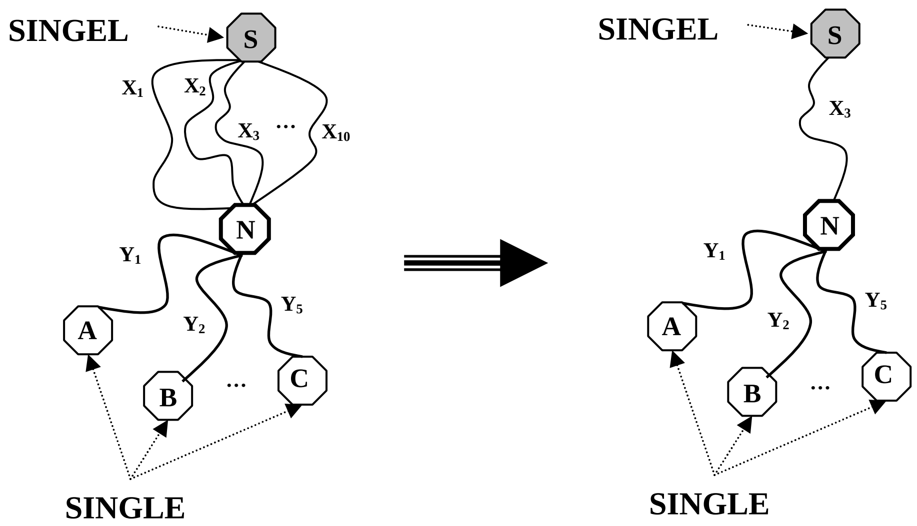
Her er det en date ekstra, og generelt vil vi alltid få en date ekstra hvis vi klarer å lage en slik vei som beskrevet. Alle datene som er markert her er gyldige, for vi beveger oss alltid annenhver gang mellom to som enten allerede er satt opp som datepartnere, eller to som har markert hverandre som potensielle datepartnere.

Et tre av forbedringsveier

Vi kan forsøke bygge ut alle slike veier fra startgutten og forsøke å ende opp hos en singel jente, klarer vi dette så kan vi gå videre å gjøre det samme med den neste single gutten. Ett problem er nå at vi ikke kan forsøke å finne absolutt alle unike forbedringsveier fra den aktuelle startgutten slik vi beskrev, for i verste tilfelle er det eksponesielt mange av dem. Det vi må innse er at det slutten av veiene som er interessant, nemlig om vi ender opp hos en jente som er ledig, ikke hvilken vei vi faktisk tok for å komme dit.

Hvis to veier som starter hos startgutten går hver sin sti og på et tidspunkt møtes i en felles node, så er det bare vits i å forbedre en av de, for vi vil uansett ha lik mulighet til å ende opp hos en singel jente. Generelt kan vi si at det aldri er noe poeng å forbedre flere stier som tar forskjellige ruter til en felles node på veien.

For eksempel kan vi anta at det er 10 stier som er forskjellig før de møtes i en node S, og vi vet at fra S kan vi bygge ut 5 forskjellige stier. Istedenfor å bygge ut alle 50 forskjellige stier fra start til slutt, så kan vi heller bare ta for oss 5 stier, som da alle ser identiske ut før S, og som forgrener seg på 5 forskjellige måter herfra.

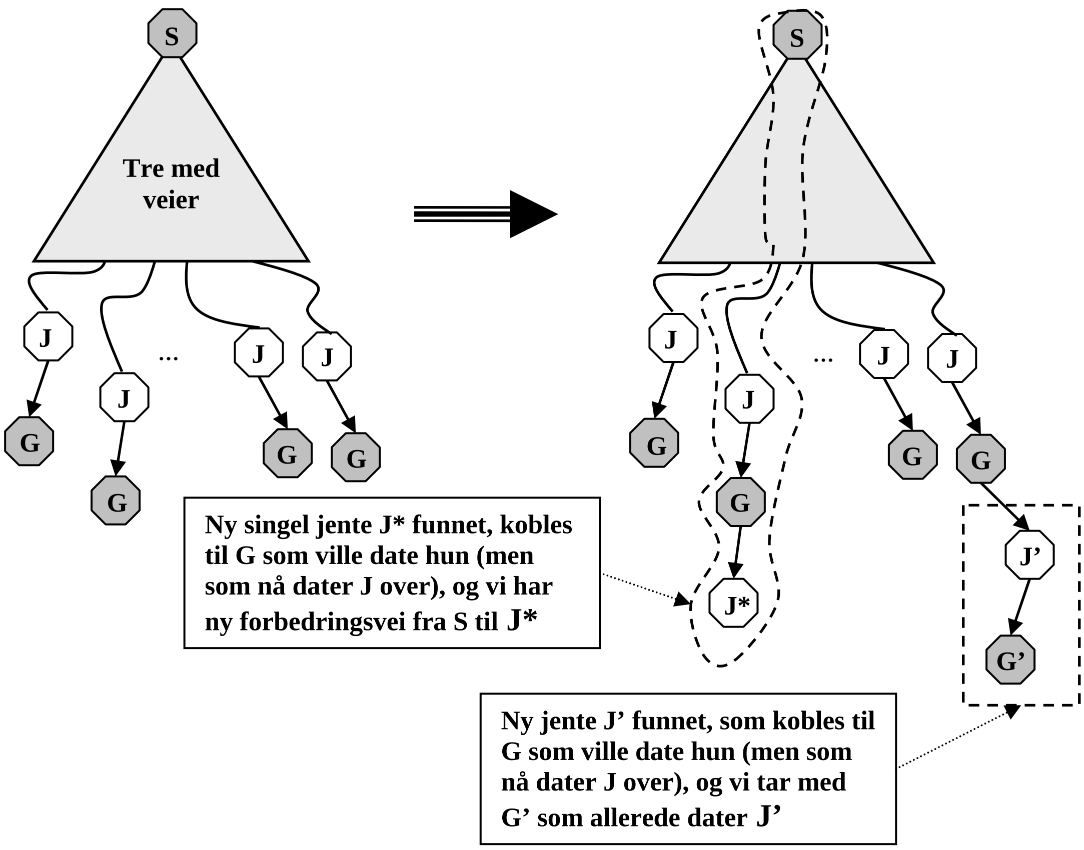


Som vi kan se fra figuren er det 10 forskjellige veier som ender opp i enhver node som er singel, for eksempel vil X_1Y_1, X_2Y_1, … , $X_{10}Y_1$ alle ende opp i A. Uavhengig av hvilken vei som velges vil man utvide koblingen med en enkelt date, derfor kan vi like gjerne se på situasjonen hvor vi har en og bare en vei til hver node A, B og C. Vi ser at vi kun trenger å finne de veiene som starter overlappende og etter hvert forgrener seg, med andre ord danner de et tre. Roten av treet er startgutten, herfra gror det lange stier som er forsøk på koblingsforbedrende veier, og når det er en forgrening i treet er det i virkeligheten to forskjellige stier som inntil nå har vært like.

Bygge trær med forbedringsveier

Treet starter først som bare en rotnode som da er startgutten. For å utvide treet velger vi en jente som ikke er i treet, men som kan nås direkte fra en gutt som er i treet. Hvis hun allerede har en date, så hekter vi hun og hennes date på under den gutten som kunne nå henne fra treet. Legg merke til at gutten som hun har en date med ikke allerede kan befinne seg i noen annen forbedringsvei i treet, for hvis han gjorde det så betyr det at han ble lagt til treet med en datepartner som han allerede hadde fra før av, og ingen kan ha to datepartnere.

Hvis hun ikke har en date så hekter vi hun også under gutten som ønsket å date henne, og da har vi en forbedringsvei fordi vi startet med en ledig gutt, endte med en ledig jente, og har datende gutt og jente annenhver gang. Som vist kan vi da invertere pardannelsene langs veien og få en kobling som har én date mer i seg, hvis vi ikke er ferdig så prøver vi å utvide treet igjen på samme måte.



Hver gang vi velger en nyt startgutt må vi naturligvis bygge et nytt tre med han som rotnode.

Når perfekt kobling er umulig

Ett problem med den første naive metoden var at vi ikke kunne vite hvorvidt vi faktisk var ferdig med oppgaven når vi ikke klarte å lage flere dates, så la oss se på hva som skjer når vi når ikke klarer å lage flere dates uten å ha en perfekt kobling med The Hungarian.

Vi ser på hva som skjer når metoden for å finne koblingsforbedrende veier kommer frem til at det ikke er mulig. Det som skjer er at når du har prøvd å lage mange koblingsforbedrende veier, så sitter du igjen med en haug med veier i treet som du aldri klarte å få til å ende hos en ledig jente, og som dermed da ikke var koblingsforbedrende. Det man kan se da er at du sitter igjen med en masse veier som alle ender i gutter som har en date, grunnen til at alle ender med gutter er at du alltid utvider en vei ved å henge på en jente og hennes datepartner.

Husk at alle de potensielle forbedringsveiene starter med gutten som opprinnelig forsøkte å skaffe en date til, de ender med en opptatt gutt og langs veien er det gutt og jente annenhver gang. Siden vi ikke kan gjøre noen veier lenger, så betyr det at ingen av guttene kan tenke seg å date noen andre jenter enn de som er i treet, ellers så kunne vi jo ha lagt til denne jenta (og muligens hennes partner). Dermed har vi et tre som inneholder en mengde gutter, og alle jentene de kunne tenke seg å date. Fra treet kan vi se at for alle guttene i treet er det en tilsvarende jente rett over som gutten skal på date med, for utenom roten, som ikke har noen jente over seg. Det betyr at det er EN gutt flere enn det er jenter i treet, så vi har påvist en gruppe gutter som er EN større enn gruppen jenter de ønsker å date, og da går det åpenbart ikke å koble alle sammen.

Hele denne prosessen med å velge en tilfeldig gutt man vil skaffe en partner til ved å gro et tre som inneholder forbedringsveier gjentas helt til man enten har koblet sammen alle, eller man ikke klarer å bygge ut et tre lenger og da har bevis for at en full kobling ikke er mulig.

Bedeho Mender
IFI – Universitet i Oslo