

Að teikna fjölskyldur*

Guðbjörn Freyr Jónsson

Íslenskri erfðagreiningu

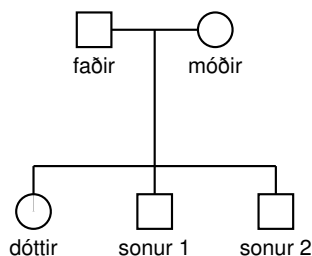
Vefútgáfa 16. desember 2004.

Ágrip Við rannsóknir á ættgengum sjúkdómum er mikilvægt að geta séð fjölskyldutengsl þeirra sem sjúkdóminn hafa. Fjallað er um reiknirit til að búa til teikningar af fjölskyldum, þar sem fylgt er þeim venjum sem skapast hafa í erfðafræðirannsóknum. Reikniritið byggir á aðferðum til að teikna lagskipt örvanet, en gerðar eru mikilvægar endurbætur sem nýta sér sérstaka eiginleika ættartjáa.

Inngangur

Myndræn framsetning upplýsinga verður æ mikilvægari eftir því sem okkur vex ásmegin í gagnaöflun og úrvinnslu. Hér verður fjallað um reiknirit til þess að teikna myndir sem sýna ættartengsl milli einstaklinga, m.ö.o. teikningar af fjölskyldum. Þetta reiknirit er notað í hugbúnaði Íslenskrar erfðagreiningar, en fyrirtækið hefur þá sérstöðu að búa yfir miklum ættfræðiupplýsingum sem notaðar eru í rannsóknum þess. Reikniritið var þróað vegna þess að sá hugbúnaður sem til var réð ekki við nema mjög takmarkaða stærð á fjölskyldum og þó svo að laga mætti teikningarnar í höndunum var mjög tímafrekt að greiða úr flækjunum. Áherslan er á notkun fjölskylduteikninga við rannsóknir á ættgengum sjúkdómum, en reikniritið nýtist almennt til að sýna ættartré.

* Þetta er endurbætt útgáfa af grein sem birtist áður í Algrím, riti útskriftarnema í tölvunarfræði við Háskóla Íslands, vorið 2002.



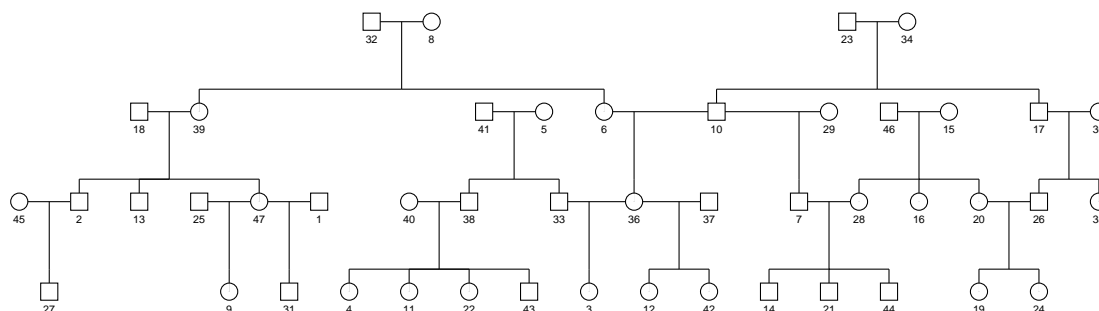
Mynd 1. Kjarnafjölskylda.

Í erfðafræðiritum er venja að teikna fjölskyldur eins og mynd 1 sýnir. Karlar eru táknaðir með kassa og konur með hringjum. Einstaklingar úr sömu kynslóð eru teiknaðir í sömu hæð á blaðinu eða á tölvuskjánum. Ef karl og kona eiga barn eða börn saman eru þau tengd saman með láréttri línu (*mökunarlínu*). Neðan úr miðri mökunarlínunni kemur lóðrétt lína sem liggur niður að annarri láréttri línu (*systkinalínu*), en frá henni koma svo aftur lóðréttar línur til barnanna. Ef barnið er aðeins eitt, er einungis lóðrétt lína frá mökunarlínunni til barnsins, en engin systkinalína.

Oft er fleiri upplýsingum komið til skila í teikningunni. Ef einstaklingur er látinn er skálína dreginn yfir kassann eða hringinn. Þeir sem hafa verið greindir með sjúkdóm eru oftast litaðir svartir, en ef sjúkdómflokkarnir eru fleiri en einn eru notaðir aðrir litir eða mynstur. Ef makar eru skyldir (innan fjölskyldunnar) er mökunarlínán höfð tvöföld. Einnig eru tvíburar tengdir saman á ákveðinn hátt, svo nokkur dæmi séu tekin. Nánari útlistun á hvernig teikna skal fjölskyldur í sjúkdómsrannsóknum má sjá í tillögu að teiknistaðli [2].

Við fyrstu sýn mætti ætla að það sé ekki svo erfitt að búa til svona teikningar, en ýmis atriði flækja málið verulega:

- Frændsystkin geta átt börn saman og þannig myndað hring í teikningunni.
- Fólk getur átt börn með fleiri en einum maka.
- Makar geta tilheyrt mismunandi kynslóðum.
- Margir í systkinahóp eiga maka sem einnig hafa foreldra á teikningunni.



Mynd 2. Stærri og flóknari fjölskylda.

- Fjölskylda er samsett úr nokkrum smærri fjölskyldum með mörgum tengingum sín á milli.

Mynd 2 sýnir fjölskyldu með 47 einstaklingum þar sem sum þessara atriða koma fram. Þar má sjá einn karl og tvær konur sem hvert um sig eiga börn með tveimur ólíkum mökum. Einnig eru þar systur (númer 20 og 28) sem eiga börn með bræðrasönnum.

Markmiðið er auðvitað að gera teikninguna eins læsilega og mögulegt er. Tores og Barillot [7] skilgreina reglur fyrir það sem þeir kalla fullkomlega teiknanlega fjölskyldu, en það er fjölskylda sem hægt er að teikna þannig að

- einstaklingarnir (þ.e. kassarnir og hringirnir) skarast ekki,
- makar séu hlið við hlið,
- systkin séu hlið við hlið en þó mega foreldralausir makar koma á milli þeirra,
- foreldrar séu beint ofan við börnin (nánar tiltekið að teikna megi lóðréttu línu frá mökunarlínu til syskinalínu) og
- línurnar skerast ekki.

Þeir sýna svo að það megi þýða þessi skilyrði yfir í klemmuvandamál fyrir bilanet (e. interval graph sandwich problem). Net er bilanet ef hver hnútur samsvarar rauntalnabili og hnútar tengjast ef bilin skarast. Klemmuvandamálið felst í því að finna slíkt net sem inniheldur ákveðið mengi leggja, en má ekki hafa leggi úr öðru mengi.

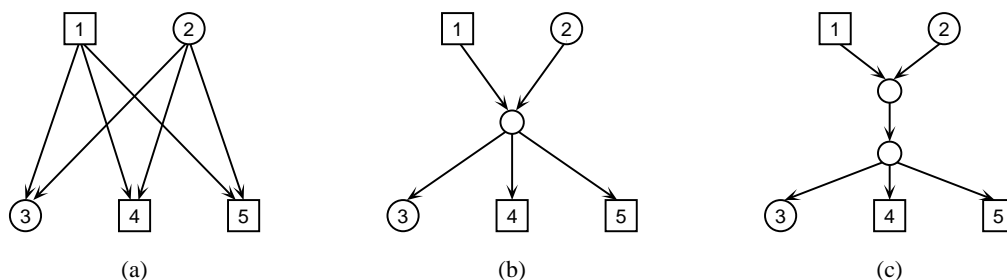
Af framansögðu ætti hins vegar að vera ljóst að ekki eru allar fjölskyldur fullkomlega teiknanlegar. Til dæmis er ekki alltaf hægt að komast hjá því að línur skerast eða að krefjast þess að mökunarlínan sé lárétt. (Ef að er gáð brýtur teikningin á mynd 2 regluna um að systkin eigi að vera hlið við hlið.) Það dugar ekki

að geta bara teiknað nógu þægilegar fjölskyldur og við viljum því aðferð sem gengur fyrir allar fjölskyldur og reynir að fylgja áður nefndum reglum eins vel og hægt er.

Grunnhugmyndin er að líta á ættartré sem örvanet og nýta sér aðferðir til að teikna slík net í lögum eða hæðum. Örvanet (V, E) er mengi af hnútum, $V = \{v_1, \dots, v_n\}$, og mengi E af örvum eða leggjum sem benda frá einum hnúti til annars. Leggjur sem bendir frá v_i til v_j er táknaður með tvenndinni (v_i, v_j) . Ættartré hafa augljósa stefnu þar sem forfeður og formæður koma á undan afkomendum sínum. Í raun eru þau órásuð örvanet (e. directed acyclic graphs) þar sem enginn getur verið sinn eigin forfaðir eða formóðir.

Það er hægt að lýsa fjölskyldutengslum með örvaneti á fleiri en einn hátt. Mynd 3 sýnir þrjár mismunandi framsetningar. Sú fyrsta hefur hnúta fyrir einstaklingana og leggi sem benda frá foreldri til barns. Við gætum auðvitað snúið þessu við og látið börnin benda á foreldrana og þannig eru upplýsingarnar yfirleitt geymdar, þ.e.a.s. fyrir hvern einstakling er skráð hverjir foreldrar hans eru. Hér höfum við hins vegar valið að láta örvanar stefna niður ættartréð. Önnur framsetningin hefur foreldra benda á mökunarhnút sem síðan bendir á sameiginleg börn þeirra. Þetta er mun nær því hvernig fjölskylda er teiknuð, en það er hægt að ganga enn lengra og hafa bæði mökunarhnút og systkinahnút. Þriðja framsetningin er einmitt þannig og með henni má lýsa nákvæmlega (með því að úhluta öllum hnútunum hnit) hvernig fjölskylda er teiknuð, eins og sést með samanburði við mynd 1.

Reikniritið sem kynnt er hér hefur, eins og flestar aðferðir sem búa til lagskiptar teikningar af örvanetum, eftirfarandi skref:



Mynd 3. Mismunandi framsetningar á fjölskyldu í formi örvanets, (a) með leggjum beint til barna, (b) með mökunarhnút og (c) með bæði mökunar- og systkinahnút.

1. Einstaklingum er skipt í kynslóðir (sem samsvarar lagskiptingu hnúta í örvaneti).
2. Þeim er raðað innan hverrar kynslóðar til að lágmarka þann fjölda lína sem skerast.
3. Hnit allra hnútanna eru ákveðin og línurnar teiknaðar.

Fyrstu tvö skrefin byggja á teikniáðferð fyrir örvanet eftir þá Gansner, Koutsofios, North og Vo [5]. Þetta er áðferðin á bak við forritið `dot` sem er hluti af `Graphviz` pakkanum frá AT&T og er hægt að fá án endurgjalds á vefnum [1]. Þetta er kannski ekki sérlega frumlegt þar sem fleiri teikniáðferðir fyrir ættartré byggja á sömu hugmynd. Til dæmis er forritið `dot` notað beint í [8], en sú útfærsla virkar ekki vel vegna þess munar sem er á teikningum af fjölskyldum og almennum örvanetum. Það skiptir nefnilega miklu máli að nýta sér sérstaka eiginleika fjölskylduneta, bæði í röðunarskrefinu og við úthlutun hnita.

Í reikniriti Gansners et al. [5] er skref 1 leyst með heiltölubestun og skref 2 með ítrun sem reynir að fækka leggjum sem skerast (leggja í kross). Við notum skref 1 óbreytt, en á skrefi 2 eru gerðar endurbætur sem nýta sér eiginleika örvaneta fyrir ættartré. Þær eru helstar að einfalda netið og vinna með nokkurs konar beinagrind af fjölskyldunni, og að byrja á forröðun sem greinir sundur megingreinar fjölskyldunnar. Í skrefi 3 er byrjað neðst og hnútum í hverri kynslóð úthlutað hnitum, en eftir hverja kynslóð eru kynslóðirnar fyrir neðan hana lagaðar til ef þurfa þykir. Í skrefum 1 og 3 notum við örvanet með bæði mökunar- og systkinahnútum (mynd 3(c)), en í skrefi 2 er notað net sem hefur mökunarhnúta en ekki systkinahnúta (mynd 3(b)).

1. Ákvörðun á kynslóðum

Fyrsta skrefið í reikniritinu er að skipta fjölskyldunni í kynslóðir. Látum (V, E) vera net fyrir fjölskylduna með bæði mökunar- og systkinahnútum (sjá mynd 3(c)). Hnútunum er raðað í kynslóðir með því að úthluta hverjum þeirra kynslóðatölu λ . Yngri kynslóðir fá hærri kynslóðatölu en þær eldri, svo við setjum skorður

$$\lambda(v) - \lambda(u) \geq 1, \quad (1)$$

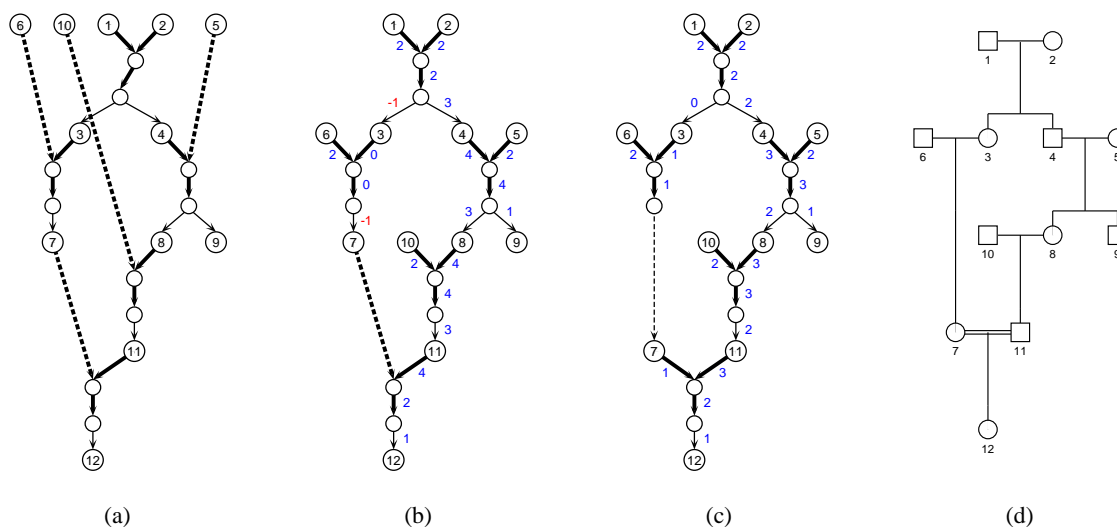
fyrir alla leggi $(u, v) \in E$. Orðið kynslóð er hér bæði notað í sinni eiginlegu merkingu (innan fjölskyldunnar) og yfir hvert lag eða hæð í örvanetinu, en þetta ætti ekki að valda misskilningi þar sem skýr samsvörun er þarna á milli.

Við úthlutum kynslóðatölum með því að lágmarka

$$\sum_{(u,v) \in E} \omega(u, v) (\lambda(v) - \lambda(u)), \quad (2)$$

að uppfylltum skorðunum (1). Hér eru $\omega(u, v)$ vægi á leggjum sem gefa okkur stjórn á hvaða leggjum er teygt. Meira vægi þýðir að það er síður teygt á þeim legg en þeim sem hafa minna vægi. Í okkar tilfalli fá leggir frá systkinahnútum til einstaklingshnúta vægi 1, en leggir frá einstaklingshnútum til mökunarhnúta og frá mökunarhnútum til systkinahnúta vægi 2.

Þetta línulega bestunarvandamál má leysa á ýmsa vegu, en við notum “nethyrnuáðferð” (e. network simplex method) Gansners og félaga [5], þó svo að ekki sé víst að hún hafi margliðuflækjustig. Áðferðin er einföld í forritun og finnur bestu lausn yfirleitt í örfáum skrefum. Reyndar má líta á þetta sem nykrað vandamál (e. dual problem) fyrir lágmarkun kostnaðs á flæði gegnum örvanet (e. minimum cost flow problem) og er áðferðin sem notuð er hér í raun nykruð útgáfa



Mynd 4. Röðun á fjölskyldu í kynslóðir. Feitir leggir hafa vægi 2 en grannir vægi 1. Brotalínur sýna leggi sem ekki eru þéttir. Í skrefi (b) eru tveir leggir með neikvæða klippitölu, en enginn í skrefi (c).

af nethyrnureikniritinu eins og því er til dæmis lýst í [3]. Þar sem lágmarkslengd allra leggja er heiltala, þá má sýna fram á að til sé besta lausn sem er heiltölu-lausn (sjá t.d. [3]). Aðferðinni verður nú lýst í grófum dráttum.

Við segjum að kynslóðaskipting sé *gjaldgeng* (e. feasible) ef $\lambda(v) - \lambda(u) \geq 1$ fyrir alla leggi (u, v) . *Slaki* leggjar er $\lambda(v) - \lambda(u) - 1$ og leggur er *þéttur* ef slaki hans er núll. Ef netið er tré (hefur enga hringi sem óstefnt net) er auðvelt að finna lausnina. Við einfaldlega vinnum okkur út frá einhverjum byrjunarhnút og úthlutum kynslóðatölum þannig að allir leggir verði þéttir. Á þessari athugun byggist aðferðin í almenna tilfellinu.

Fyrst er búin til gjaldgeng kynslóðaskipting (mynd 4(a)). Þá er netinu skipt í samhengisþætti byggða á þéttum leggjum og búið til tré sem spannar stærsta samhengisþáttinn og notar aðeins þetta leggi. Tréð er stækkað með því að bæta við legg frá trjáhnúti til hnúts utan trésins með minnstum slaka og sá leggur er jafnframt gerður þéttur. Þetta tryggir að lausnin er áfram gjaldgeng og þannig er tréð látið vaxa þar til það nær yfir allt netið (mynd 4(b)).

Ef leggur er fjarlægður úr trénu skiptist það í tvo hluta, haushluta og sporðhluta, sem miðast við haus og sporð leggjarins sem var tekinn burt. Skilgreinum *klippitölu* leggjarins sem summu vægja þeirra leggja í netinu sem benda frá sporðhluta til haushluta mín-

us summu vægja leggjanna sem benda frá haushluta til sporðhluta. Klippitölur allra trjáleggja eru sýndar á mynd 4(b). Neikvæð klippitala bendir til þess að lækka megi (2) með því að lengja legginn eins mikið og hægt er (ef það er hægt á annað borð), eða þar til einhver leggur frá haushluta til sporðhluta verður þéttur (sjá skilyrði fyrir lágmarkslausn í línulegri bestun [3]).

Ef engin klippitala er neikvæð, þá höfum við gjaldgenga kynslóðaskiptingu sem lágmarkar (2). Annars er einn leggja sem hafa neikvæða klippitölu fjarlægður úr trénu og í staðinn tekinn leggur frá haushluta til sporðhluta með minnstum slaka (til að halda skiptingunni gjaldgengi). Um leið eru kynslóðatölum annars hlutans breytt svo að nýi leggurinn verði þéttur. Þetta er endurtekið þar til enginn leggur hefur neikvæða klippitölu (mynd 4(c)). Það er að vísu ekki tryggt að fallið (2) lækki í hverju skrefi og jafnvel hugsanlegt að reikniritið fari í óendanlega lykkju.¹ Það má koma í veg fyrir óendanlegar lykkjur með vali á leggjum til að fjarlægja og bæta við, ef margir koma til greina. Þetta er hins vegar svo fátítt að ekki hefur reynst þörf fyrir slíkar ráðstafanir hér frekar en í [5].

Að lokum er öllum kynslóðatölunum hliðrað þannig að lágsta talan verði núll. Þá höfum við raðað

¹ Þetta gildir almennt um hyrnureikniritið (e. simplex algorithm) og má lesa um í mörgum bókum um línulega bestun, eins og t.d. [3].

öllum hnútunum í kynslóðir 0, 1, 2, o.s.frv. Ef netið hefur leggi sem spanna meira en eitt kynslóðabil, þá er þeim skipt upp og innleiddir sýndarhnútar þannig að hver leggur bendi frá einni kynslóð til næstu kynslóðar fyrir neðan. Þetta er nauðsynlegt fyrir röðunarferlið sem verður lýst næst.

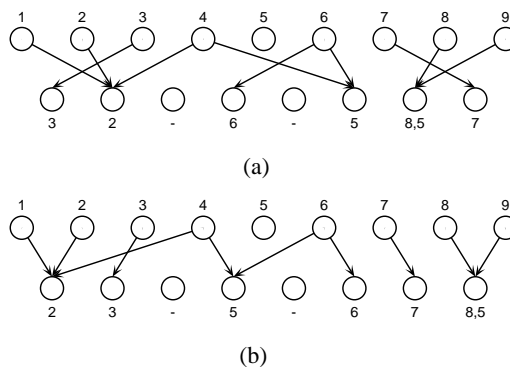
2. Röðun innan kynslóða

Þegar búið er að skipta hnútunum í kynslóðir, reynum við að raða þeim innan hvorrar kynslóðar þannig að það verði sem fæstar *krossanir* á leggjum, en með krossun er átt við að tveir leggir skerist (liggi í kross). Þetta er erfiðasta og tímafrekasta, en jafnframt mikilvægasta skrefið í reikniritinu. Hversu læsileg teikningin er veltur mikið á því að ekki sé of mikið af línunum þvers og kruss yfir blaðið. Að finna þá röðun sem gefur fæstar krossanir er NP-erfið vandamál, jafnvel þó aðeins tvö lög séu í örvanetinu [4]. Því er notuð reikniaðferð sem reynir að komast nálægt lágmarki án þess að vera of tímafrek. Þar sem við viljum að mökunarhnútar séu beint fyrir ofan samsvarandi systkinahnúta, fellum við þessa hnúta saman, þ.e.a.s. við notum netaframsetninguna á mynd 3(b).

2.1. Ítrun á veginni miðgildisröðun

Að grunni til byggir röðunin á þeirri röðunaraðferð sem notuð er í forritinu dot [5]. Þar er hnútunum í hverju lagi fyrst raðað af handahófi. Svo er framkvæmd röð ítrana þar sem farið er til skiptis upp og niður netið. Á niðurleið er hverju lagi raðað miðað við næsta lag fyrir ofan, en á uppleið er miðað við næsta lag fyrir neðan. Hnútunum er raðað eftir vegnu miðgildi staðsetningu þeirra hnúta sem þeir tengjast í viðmiðunarlaginu. Til eru ýmsar aðrar útfærslur á þessu skrefi, eins og til dæmis að nota meðalgildi [6].

Segjum að við séum á niðurleið og að við höfum raðað lagi númer k . Hnútunum í lagi k eru gefnar tölur $1, 2, 3, \dots$ eftir því hvar þeir eru í röðinni. Skoðum hnút v í kynslóð $k + 1$ og gerum ráð fyrir að hann tengist hnútum með tölugildi $p_1 < \dots < p_m$ í kynslóð k . Ef m er oddatala, $m = 2r - 1$, þá fær v miðgildið p_r . Ef m er slétt tala, $m = 2r$, þá fær hnúturinn v tölugildið $(\beta p_r + \alpha p_{r+1}) / (\alpha + \beta)$, þar sem $\alpha = p_r - p_1$ og $\beta = p_m - p_{r+1}$. (Ef $r = 1$, þá setjum við $\alpha = \beta = 1$.) Síðan eru hnútunum í lagi $k + 1$ raðað eftir þessum tölugildum, nema að hnútar sem tengjast engum í lagi k eru látin vera óhreyfðir. Mynd 5 sýnir dæmi um svona röðunarskref.



Mynd 5. Röðunarskref sem fækkar krossunum úr 5 í 1.

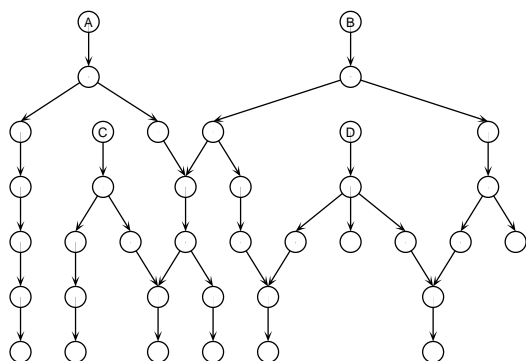
Einnig er í hverri ítrun farið í gegnum netið og hnútum hlið við hlið víxlað ef það fækkar krossunum. Í lok hvorrar ítrunar er athugað hvort tekist hefur að fækka krossunum og ef svo er er nýja röðunin geymd en annars er haldið áfram með þá gömlu. Ef engri framþróun er náð tvisvar í röð (þ.e. bæði á uppleið og niðurleið) er ítrunum hætt.

2.2. Einföldun netsins

Hingað til hafa eingöngu verið notaðar aðferðir úr [5] sem virka almennt á örvanet. En fjölskylduörvanet eins og þau sem við fjöllum um hér hafa ýmsa sérstaka eiginleika sem má nýta til þess að einfalda röðunina. Til dæmis hafa ættartré oft greinar af afkomendum sem eiga aðeins foreldralausa maka (þ.e.a.s. foreldrarnir eru ekki hafðir með). Slíkar greinar er einfalt að teikna og því óþarfi að hafa þær í heilu lagi inni í röðunarferlinu. Það sama gildir um greinar þar sem eingöngu eru forfeður og formæður einhvers ákveðins einstaklings og engin skörun er á milli ættleggja.

Hugmyndin er að í stað þess að raða hnútunum í upphaflega netinu, sé hnútunum einfaldara nets raðað, þar sem þetta einfaldara net er einhvers konar beinagrind af fjölskyldunni. Þessi beinagrind fæst með því að fyrir áður nefndar ættgreinar eru hnútar í sama lagi felldir saman í einn. Einnig eru foreldralausir einstaklingar felldir saman við maka sinn. Mynd 6 sýnir einfaldað net fyrir fjölskylduna sem teiknuð er á mynd 2.

Þegar einfaldaða netinu hefur verið raðað eru þeir hnútar sem voru sameinaðir slitnir aftur í sundur. Þeim er raðað þannig að áður nefndar ættgreinar hafi engar krossanir og settir inn meginröðina í stað sameinaða hnútsins. Einnig þarf að gæta þess að makar raðist rétt



Mynd 6. Einfaldað net fyrir fjölskylduna úr mynd 2.

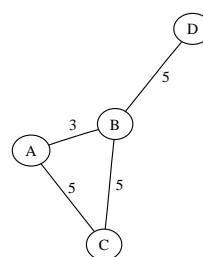
upp og huga að ýmsum öðrum smáatriðum sem ekki verður farið út í hér.

2.3. Forröðun

Röðunarferlið, eins og því hefur verið lýst, virkar að mörgu leyti vel. Gallinn við það er sá að það á erfitt með stórar breytingar, þannig að ef það villist af leið eða ef upphafsroðunin er slæm, næst oft ekki að greiða úr flækjunum. Þetta gerist gjarnan með stærri fjölskyldur sem hafa margar samtengdar ættgreinar. Því er byrjað á grófri röðun sem reynir að greiða í sundur meginættgreinar fjölskyldunnar.

Skoðum alla hnúta sem hafa enga forfeður, þ.e. engin ör bendir á þá. Búum til net með þessum hnútum þar sem við setjum legg milli tveggja hnúta ef þeir hafa sameiginlega afkomendur. Leggurinn fær lengd eftir því í hvaða kynslóð fyrstu sameiginlegu afkomendurnir eru. Hugmyndin er svo að raða hnútum þessa nets (í einfalda röð) eftir fjarlægð, þannig að hnútar tengdir með stuttum legg raðist nálægt hvor öðrum. Það má vissulega hugsa sér ýmsar leiðir til að finna slíka röðun. Til dæmis mætti setja þetta fram sem bestunarvandamál og leysa það. Í stað þess finnum við einfaldlega þá tvo hnúta sem eru lengst hvor frá öðrum í þessu nýja neti og röðum hnútunum í netinu eftir mismun fjarlægðanna frá þessum tveimur hnútum.

Tökum netið á mynd 6 sem dæmi. Þar myndum við búa til net með hnútunum sem merktir eru A , B , C og D . Það væri leggur milli A og B af lengd 3 og leggir milli A og C , B og C , og B og D af lengd 5 (sjá mynd 7). Þá eru C og D lengst hvor frá öðrum og hnútarnir fá röðina $C - A - B - D$.



C	A	B	D
0	5	5	10
10	8	5	0
-10	-3	0	10

Mynd 7. Forfeðranet fyrir netið úr mynd 6 og röðun hnútanna eftir mismun fjarlægða frá C og D .

Þar með höfum við röð á forfeðralausum hnútum. Þessa röð notum við síðan til að raða öllum hnútunum í upphaflega örvanetinu. Í dæminu fá hnútarnir C , A , B , D tölugildi 1, 2, 3, 4, hver hnútur fær svo meðalgildi þeirra sem benda á hann og síðan er hverju lagi raðað eftir þessum tölum.

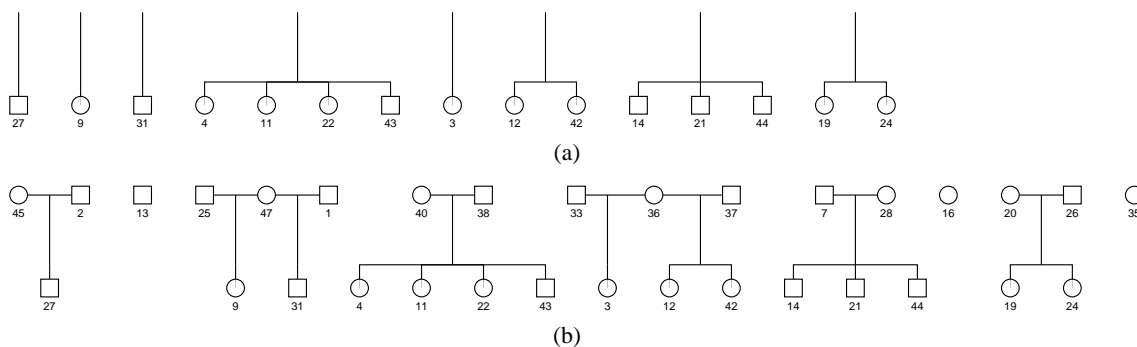
Það má gera þetta enn almennar og taka alla hnúta í ákveðnu lagi (t.d. því sem hefur flesta hnúta), alla forfeðralausa hnúta fyrir neðan þetta lag og alla afkomendalaus hnúta fyrir ofan lagið. Fjarlægðir eru reiknaðar með því að telja frá þessu lagi niður á við og halda svo áfram efst uppi. Þegar öllum þessum hnútum hefur verið raðað innbyrðis er hinum gefin tölugildi með því að rekja sig bæði upp og niður frá laginu. Þetta er það sem gert er í reikniritinu og reyndar prófað fyrir öll lög af einstaklingshnútum og síðan er besta röðunin valin.

3. Úthlutun hnita

Lokaskrefið er að úthluta öllum hnútunum hnit. Við skilgreinum bil Δy milli kynslóða og lágmarksbil Δx milli einstaklinga. Systkinahnútar og þar með systkinalínur eru teiknaðar $\Delta y/4$ fyrir ofan samsvarandi systkinahóp.

Hnitaúthlutunin byrjar neðst og rekur sig svo upp eftir fjölskyldunetinu. Einstaklingunum í neðstu kynslóðinni er raðað í hæð $y = \Delta y$ með jöfnu millibili: Δx , $2\Delta x$, $3\Delta x$, o.s.frv. Síðan fá systkinahnútar og samsvarandi mökunarhnútar x -hnit sem er mitt á milli þeirra barna sem eru lengst til vinstri og lengst til hægri í barnahópnum (sjá mynd 8(a)).

Þá er einstaklingunum í næstu kynslóð fyrir ofan komið fyrir á milli mökunarhnútanna (sjá mynd 8(b)). Ef það er nóg pláss er þeim einfaldlega raðað með jöfnu millibili á milli mökunarhnútanna, en þó þannig



Mynd 8. Úthlutun hnita.

að fjarlægðin frá mökunarhnút til maka sé $\Delta x/2$. Ef það er hins vegar ekki nóg pláss á milli mökunarhnúta, þá er öllum mökunarhnútunum til hægri hliðrað lengra til hægri svo hægt sé að koma fyrir einstaklingshnútunum með millibili Δx .

Næst staðsetjum við systkinahnútana fyrir þessa kynslóð eins og fyrir þá neðstu og þannig er haldið áfram upp ættartréð. Í hvert skipti sem einstaklingum í ákveðinni kynslóð hefur verið komið fyrir, þá þarf að hliðra öllum í kynslóðunum fyrir neðan í samræmi við hliðranirnar á mökunarhnútunum. Þetta er gert með því að hliðra systkinahóp og mökum þeirra um sömu fjarlægð og samsvarandi mökunarhnútur. Einstaklingum á milli slíkra hópa er hliðrað um meðaltal hliðrananna á hópunum til sinnar hvorrar hliðar.

Hnitaúthlutunin, eins og henni hefur verið lýst hér, virkar vel ef tekist hefur að raða hnútunum innan kynslóða þannig að engir leggir skerist. Ef það hefur hins vegar ekki tekist, þá verður t.d. mun erfiðara að ákveða hversu mikið eigi að hliðra einstaklingum í neðri kynslóðum. Þannig að í reynd er aðferðin sem er notuð nokkuð flóknari, þó grunnhugmyndin sé sú sama.

Lokaorð

Hér hefur verið stiklað á stóru í gegnum þá aðferð sem nú er notuð við að ákvarða hnit þegar fjölskyldur eru teiknaðar í hugbúnaði Íslenskrar erfðagreiningar. Leitast hefur verið við að koma meginhugmyndunum til skila í stað þess að lýsa öllum atriðum nákvæmlega. Í aðferðinni felst veruleg framför frá þeim teikniforritum sem áður voru notuð, þó enn megi margt bæta. Það er mest pláss fyrir endurbætur í síðasta skrefinu, sérstaklega þegar fjölskyldan er það flókin að ekki næst að koma í veg fyrir krossanir á leggjum.

Ekki hefur verið gerður neinn formlegur samanburður við önnur teikniforrit, enda ekki augljóst hvernig mæla eigi gæði teikninga. Höfundur hefur þó ekki fundið neitt forrit sem stenst samanburð þegar kemur að sémilega stórum fjölskyldum (með yfir hundrað einstaklinga). Áherslan við þróun forritsins var meiri á gæði teikninganna en hraða og því má eflaust gera forritið hraðvirkara. Það tekur þó ekki nema nokkrar sekúndur að teikna dæmigerða þrjú hundruð manna fjölskyldu. Tíminn getur vaxið nokkuð hratt með stærð fjölskylda, vegna þess að líkurnar á flækjum aukast og erfitt getur verið að gera snyrtilega teikningu.

Höfundur er mjög þakklátur fyrir að fá svo skemmtilegt verkefni sem þróun þessa reikniritis hefur verið og vill einnig þakka Kristjáni Jónassyni sérstaklega fyrir allar þær góðu ábendingar um endurbætur sem hann kom með. Þess má að lokum geta að margar af þeim myndum sem fylgja greininni voru búnar til af teikniforritinu sjálfu, þ.m.t. myndir af örvanetum enda er reikniritið í grunninn gert fyrir örvanet. Mynd 7 var hins vegar teiknuð með forritinu `neato` sem líkt og `dot` er hluti af `Graphviz` [1], en teiknar óstefnd net.

Summary: When researching genetic diseases it is important to be able to visualize the family relations of the subjects. This article describes an algorithm for drawing pedigrees automatically, in a fashion customary in genetic research. Creating a readable drawing can be hard if there are many interrelations between the members of the pedigree, but the algorithm presented here performs quite well on moderately sized families. It is based on a heuristic for hierarchical drawings of directed graphs, but with a number of improvements utilizing the special properties of pedigrees.

Heimildir

- [1] AT&T Labs. Graphviz – tekniforrit fyrir net. www.research.att.com/sw/tools/graphviz/.
- [2] R. L. Bennett, K. A. Steinhaus, S. B. Uhrich, C. K. O’Sullivan, R. G. Resta, D. Lochner-Doyle, D. S. Markel, V. Vincent og J. Hamanishi. Recommendations for standardized human pedigree nomenclature. *Am. J. Hum. Genet.*, **56**, 1995, 745–752.
- [3] G. B. Dantzig og M. N. Thapa. *Linear Programming 1: Introduction*. Springer-Verlag New York, 1997.
- [4] P. Eades, B. McKay og N. Wormald. On an edge crossing problem. In *Proc. 9th Australian Computer Science Conference*, Australian National University, 1986, 327–334.
- [5] E. R. Gansner, E. Koutsofios, S. C. North og K.-P. Vo. A technique for drawing directed graphs. *IEEE Transactions on Software Engineering*, **19**(3), 1993, 214–230.
- [6] K. Sugiyama, S. Tagawa og M. Toda. Methods for visual understanding of hierarchical system structures. *IEEE Transactions on Systems, Man, and Cybernetics*, bf 11(2), 1981, 109–125.
- [7] F. Tores og E. Barillot. The art of pedigree drawing: algorithmic aspects. *Bioinformatics*, **17**(2), 2001, 174–179.
- [8] L. Wong. Visualization and manipulation of pedigree diagrams. *Genome Informatics*, **11**, 2000, 63–72.

Um höfundinn: Guðbjörn er fæddur árið 1969. Hann útskrifaðist frá Menntaskólanum á Akureyri 1989, lauk B.S. prófi í stærðfræði frá Háskóla Íslands 1992 og M.S. í stærðfræði frá University of Illinois, Urbana-Champaign, 1995. Guðbjörn lauk doktorsprófi í hagnýtri stærðfræði frá Cornell University 2000 og hefur starfað síðan hjá Íslenskri erfðagreiningu.

Íslensk erfðagreining
Sturlugötu 8
IS-101 Reykjavík
gfj@decode.is

Móttekin: 9. júlí 2004