

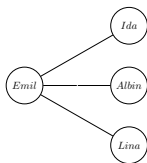
# Nyttan av att upptäcka mönster i stora nätverk

POPULÄRVETENSKAPLIG SAMMANFATTNING **David Larsson, Antonina Tokarchuk**

Att hitta och undersöka mönster i nätverk har stora användningsområden inom biologi, beteendevetenskap och smart marknadsföring. Vissa sätt att göra detta är snabbare än andra och kan leda till förbättringar inom ovanstående områden.

## Introduktion

Anta att Emil har tre kompisar: Albin, Ida och Lina, som inte känner varandra. På bilden nedan kan ni se personerna som små cirklar och deras vänskaper som streck mellan cirklarna.



Detta sätt att representera samband kallas en graf. En graf är ett väldigt generellt och mycket praktiskt sätt att representera data. Den består av noder och kopplingar (som kallas kanter) mellan dem. I många sammanhang kan man strukturera upp information genom att representera den som en graf. Ett exempel kan vara en karta, där varje stad är en nod och vägarna mellan dem är kanter. Internet kan också ses som en graf, där varje webbsida är en nod. Om en webbsida innehåller en länk till en annan webbsida, så finns det en kant mellan dem.

## Tillämpningar

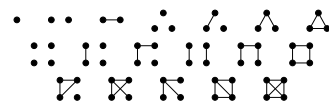
Grafer som på bilden ovan kallas sociala nätverk och studeras inom antropologi och beteendevetenskap. Till exempel, om Albin, Ida och Lina har oberoende av varandra börjat köpa ekologisk mjölk, så är Emil mycket benägen att också göra det. Sådana trender kan utnyttjas inom smart marknadsföring. Detta är företag såsom Facebook mycket intresserade av.

Inom biologi och medicin är det viktigt att veta hur proteiner i kroppen samverkar, för att förstå hur vi ska kunna vårda sjuka människor och utveckla bättre mediciner. Ett sätt att skapa sig en bild av hur proteiner fungerar

er ihop är just att göra grafer av dem. Varje protein blir en nod i nätverket och varje kant motsvarar att de två sammanknutna proteinerna interagerar på något vis. Även här är det av intresse att upptäcka olika mönster i det stora nätverk som proteinernas interagerande bygger upp. På så vis kan man förutspå till exempel biverkningar.

## Resultat

I vårt arbete har vi studerat olika algoritmer för att avgöra om en graf innehåller ett specifikt mönster (subgraf) och undersökt vilka av dem som är snabbast. De subgrafer som vi har betraktat består av upp till 4 noder. Totalt finns det 18 sådana subgrafer, vilka visas i figuren nedan.



Alla möjliga subgrafer innehållande upp till fyra noder

Generellt sett har vi jämfört två sorters algoritmer: deterministiska och Monte Carlo-algoritmer. Deterministiska algoritmer ger ett garanterat korrekt resultat medan Monte Carlo-algoritmerna get ett korrekt resultat med en viss sannolikhet. Ju fler gånger man kör dem, desto säkrare blir resultatet. I många fall visar det sig att Monte Carlo-algoritmerna är snabbare.

Dessutom har vi undersökt en teknik som kallas ordparallelism. Den gör att man exempelvis kan multiplicera matriser mycket snabbare än vanligt. Eftersom grafer kan även representeras som matriser, är denna teknik högst tillämplig. Detta beskriver vi noggrant i rapporten. Framöver hoppas vi att någon vidareutvecklar vårt arbete och tar det till nästa nivå, för att öka förståelsen för hur nätverk fungerar.