

Hauptseminar „Algorithmen und Komplexität“
Wintersemester 2011/12

Populäre Matchings und stabile Paarungen*

Felix Neumann
Version zur Veröffentlichung

19. Januar 2012

1 Einführung

Ein in der Informatik häufig anzutreffendes Problem ist die Paarung von Objekten, bekannt als *Matching-* oder *Hochzeitsproblem*. Reale Anwendungsfälle sind etwa die Zuordnung von Konsumenten auf Produzenten, von Bewerbungen auf verfügbare Stellen oder die Zuordnung im Rahmen einer Partnervermittlung. Dabei kann es vorkommen, dass die Objekte bestimmte Paarungen bevorzugen. Als *populär* bezeichnet man die Lösung eines Matchingproblems dann, wenn keine andere Lösung existiert, welche eine Mehrheit der Objekte der populären Lösung vorziehen würden. Zu einem Matchingproblem können möglicherweise populäre Lösungen verschiedener Kardinalität existieren, d.h. es wurden eine unterschiedliche Anzahl an Objekten gepaart. Bisher waren dabei vor allem populäre Matchings minimaler Kardinalität (*stabile Matchings*) bekannt, bei welchen keine zwei Objekte existieren, die nicht miteinander gepaart sind, sich aber gegenseitig ihrer tatsächlichen Paarung bevorzugen würden. Daneben existieren jedoch auch populäre Matchings höherer Kardinalität, bei welchen derartige „instabile“ Paarungen zugunsten der Popularität der gesamten Lösung in bestimmtem Maße akzeptiert werden.

Die hier nachvollzogene Arbeit [HK11] beschäftigt sich mit der Frage, wie solche populäre Lösungen maximaler Kardinalität berechnet werden können und stellt hierzu einen effizienten Algorithmus vor. Zudem wird eine Methode aufgezeigt, um eine gegebene Lösung eines Matchingproblems in Linearzeit auf ihre Popularität hin zu testen.

*Dieser Seminarbeitrag vollzieht Sätze, Algorithmen und Beweise aus Huang, Kavitha 2011: *Popular Matchings in the Stable Marriage Problem* nach.

1.1 Problemstellung

Gegeben ist ein bipartiter Graph $G = (\mathcal{A} \cup \mathcal{B}, E)$ mit Partitionen \mathcal{A} und \mathcal{B} ; die Menge aller Knoten sei $V = \mathcal{A} \cup \mathcal{B}$, weiterhin seien $|V| = n$, $|E| = m$ und der Graph zusammenhängend. Ein Matching $M \subseteq E$ auf G ist eine Menge an Kanten, die keinen gemeinsamen Knoten teilen. Die Kanten liegen stets in $\mathcal{A} \times \mathcal{B}$, es werden nachfolgend Knoten in \mathcal{A} „Männer“, Knoten in \mathcal{B} „Frauen“ genannt. Sei $M(u)$ die Bezeichnung des Nachbarn von u in M ; ist u in M isoliert, sei $M(u) = \varepsilon$.

Zu jedem Knoten existiert eine totale Ordnung seiner Nachbarn, dies ist seine Präferenzliste¹. Ein Knoten muss also nicht alle Knoten der anderen Seite in seiner Liste aufführen; doch ist ein Knoten u in der Liste eines Knoten v , so ist v auch in der Liste von u . Ein Knoten u *bevorzugt* dabei einen Knoten x gegenüber Knoten y , wenn x gegenüber y gemäß der Präferenzliste von u stärker präferiert wird; u bevorzugt die Paarung gegenüber der Isolation. Im optimalen Fall paart das Matching M jeden Knoten mit seinem meist bevorzugten Nachbarn, dies ist jedoch nicht immer möglich. Es existieren verschiedene Ansätze ein Kriterium zu finden, nach welchem „geeignete“ Matchings beurteilt werden können.

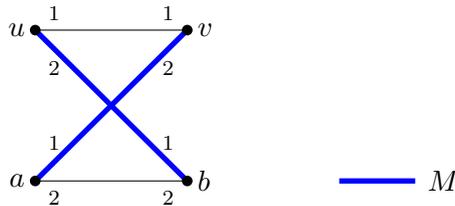


Abbildung 1: Die Kante (u, v) ist blockierend.

Ein solches Kriterium ist die *Stabilität* [GS62]. Ein stabiles Matching M vermeidet eine bestimmte Situation: nämlich dass die Knoten u und v in M nicht miteinander gepaart sind, sich jedoch gegenüber ihrer tatsächlichen Paarung in M gegenseitig bevorzugen würden (vgl. Abbildung 1). Die Kante $(u, v) \in E$ heißt nachfolgend *blockierend* bezüglich des Matchings M . Ein Matching ist also stabil, wenn es keine blockierenden Kanten erlaubt. Für einen Graph G gilt, dass alle stabilen Matchings zu diesem Graphen die selbe Knotenmenge paaren; diese wird nachfolgend U genannt.

[HK11] befasst sich mit einem anderen Kriterium, der *Popularität*. Dabei heißt das Matching M *populärer* als ein Matching M' , wenn mehr Knoten ihre Zuordnung in M gegenüber ihrer Zuordnung in M' bevorzugen als umgekehrt. Ein Matching M heißt *populär*, wenn kein populäreres Matching auf G existiert.

¹Präferenzlisten werden in Abbildungen als den Knoten und Kanten angeheftete Zahlen dargestellt; eine kleinere Zahl bedeutet stärkere Präferenz.

Es stellt sich heraus, dass die Menge der stabilen Matchings eine Teilmenge der Menge der populären Matchings ist. Zu jedem bipartiten Graph G kann ein stabiles Matching gefunden werden (hierzu geben [GS62] einen Algorithmus an), ergo existiert auch stets ein populäres Matching. Wie bereits erwähnt, sind stabile Matchings populäre Matchings minimaler Kardinalität: jedes Matching, das weniger Paarungen vornimmt, ist nicht mehr populär. Auch können populäre Matchings existieren, welche mehr Paarungen vornehmen, jedoch dann nicht mehr stabil sind.

Im Rahmen des vorliegenden Seminarbeitrags werden populäre Matchings maximaler Kardinalität gesucht. Aus dem zugrundeliegenden Paper wird ein hinreichendes Kriterium vorgestellt, anhand dessen solche Matchings identifiziert werden können, sowie ein Algorithmus, der sie effizient berechnen kann. Hierzu werden gezielt blockierende Kanten zugelassen, ohne jedoch die Mehrheit der Knoten ein anderes Matching bevorzugen zu lassen. Diese Idee wird bei [HK11] als „Stabilität im Großen“ bezeichnet.

2 Formalisierung

In diesem Abschnitt werden einige Definitionen eingeführt, die im verbleibenden Teil der Arbeit nützlich sein werden.

Zuerst wird die Präferenzliste der Knoten in einer mathematischen Funktion ausgedrückt; sie wird als Votum des Knoten für oder gegen eine Paarung modelliert.

Definition 1. *Seien $u \in V$, sowie x und y Nachbarn von u . Dann ist*

$$\text{vote}_u(x, y) = \begin{cases} 1, & \text{falls } u \text{ den Knoten } x \text{ gegenüber } y \text{ pr\u00e4feriert} \\ -1, & \text{falls } u \text{ den Knoten } y \text{ gegenüber } x \text{ pr\u00e4feriert} \\ 0, & \text{sonst} \end{cases}$$

Insbesondere ist $\forall x \in V, x \neq u : \text{vote}_u(\varepsilon, x) = -1$, das heißt jeder Knoten bevorzugt die Paarung mit einem beliebigen Knoten gegenüber der Isolation im Matching.

Anhand dessen kann nun das Konzept blockierender Kanten formal erfasst werden, indem jede Kante, die nicht Teil eines betrachteten Matchings ist, von den beteiligten Knoten bewertet wird. Diese Bewertung wird als Kantenbeschriftung aufgefasst.

Definition 2. *Sei M ein Matching auf G und $e = (u, v) \in E \setminus M$. Dann ist*

$$\text{label}_M(e) = (\text{vote}_u(v, M(u)), \text{vote}_v(u, M(v)))$$

Eine blockierende Kante e wird so mit dem Tupel $(1, 1)$ beschriftet, da beide Knoten e gegenüber ihrer Paarung in M bevorzugen.

Nun können mittels der Voten verschiedene Matchings miteinander verglichen werden, was somit den Vergleich ihrer Popularität ermöglicht.

Definition 3. Seien M_1, M_2 Matchings auf G . Dann ist

$$\Delta(M_1, M_2) = \sum_{v \in V} \text{vote}_v(M_1(v), M_2(v))$$

Definition 4. Seien M_1, M_2 Matchings auf G . M_1 heißt „populärer“ als M_2 , geschrieben $M_1 \succ M_2$, genau dann, wenn $\Delta(M_1, M_2) > 0$. M_1 heißt „populär“, wenn für kein Matching M auf G gilt, dass $M \succ M_1$.

Im Rahmen nachfolgender Sätze und Beweise werden einige zusätzliche Strukturen in Graphen benötigt.

Definition 5. Sei G ein Graph und M ein Matching auf G , so ist G_M derjenige Graph, der durch Löschung aller Kanten $e \in E \setminus M$ entsteht, für die $\text{label}_M(e) = (-1, -1)$ gilt.

Definition 6. Sei M ein Matching auf G , sowie ρ ein Pfad oder Kreis in G . ρ heißt

- alternierend in Bezug auf M , wenn abwechselnde Kanten des Pfades (nicht) zu M gehören, sowie
- augmentierend in Bezug auf M , wenn ρ ein alternierender Pfad ist, dessen Endpunkte in M isoliert sind.

Definition 7. Sei M ein Matching in G und R eine Menge von Knoten aus G . M heißt „ R -perfekt“, wenn alle Knoten in R inzident zu einer Kante in M sind.

Für zwei Mengen A und B enthält die Menge $A \oplus B$ genau diejenigen Elemente, die in genau einer der beiden Mengen A oder B enthalten sind.

3 Charakterisierung

Nachdem die Problemstellung erfasst und grundlegende Definitionen gemacht wurden, stellt der vorliegende Abschnitt nun den Ansatz von [HK11] zur Charakterisierung populärer Matchings vor, auf dem die Problemlösung basiert.

Hierzu formulieren die Autoren zuerst hinreichende und notwendige Bedingungen für die Popularität eines Matchings, und geben anschließend eine hinreichende Bedingung für populäre Matchings maximaler Kardinalität an.

Satz 1. Sei G ein Graph und M ein Matching auf G . Seien die Kanten $e \in E \setminus M$ mittels $\text{label}_M(e)$ beschriftet. M ist populär genau dann, wenn folgende Bedingungen gelten:

- (i) Es existiert kein alternierender Kreis bezüglich M in G_M , der eine mit $(1, 1)$ markierte Kante enthält.
- (ii) Es existiert kein alternierender Pfad bezüglich M in G_M , der von einem in M isolierten Knoten ausgeht und eine mit $(1, 1)$ markierte Kante enthält.
- (iii) Es existiert kein alternierender Pfad bezüglich M in G_M , der zwei mit $(1, 1)$ markierte Kanten enthält.

Beweis. Rückrichtung. Sei M' ein beliebiges Matching. Im folgenden wird angenommen, dass die Bedingungen (i) bis (iii) gelten. Zu zeigen ist, dass M populär ist, also $\Delta(M', M) \leq 0$. Betrachte also zu jedem $v \in V$ den Wert von $\text{vote}_v(M'(v), M(v))$.

Enthält M' bezüglich M eine mit $(-1, -1)$ beschriftete Kante, so würde sich das Votum der beiden Knoten nicht ändern, wenn sie in M' ungepaart wären. Daher wird nachfolgend angenommen, dass M' keine mit $(-1, -1)$ beschrifteten Kanten enthält; es kann also G_M aus Definition 5 betrachtet werden.

Knoten $v \in V$, für die $M'(v) = M(v)$ gilt, tragen nichts zur Summe der Voten bei, betrachte daher in G_M nur die Kanten in $M \oplus M'$. Zeige nun für jede Zusammenhangskomponente ρ in dem sich so ergebenden Graphen, dass

$$\sum_{v \in V(\rho)} \text{vote}_v(M'(v), M(v)) \leq 0,$$

woraus folgt, dass $\Delta(M', M) \leq 0$.

Der Maximalgrad in $M \oplus M'$ ist 2, daher sind alle Zusammenhangskomponenten darin alternierende Kreise oder alternierende Pfade. Sei ρ eine beliebige solche Komponente, unterscheide nachfolgend nach ihrer Art.

- (a) Die Komponente ist ein alternierender Kreis. Dann ist jeder Knoten mit einer Kante aus M' verbunden, und die Summe der Voten lässt sich aus den Kantenbeschriftungen bezüglich M berechnen (beschriftet sind dann genau die Kanten von M'):

$$\begin{aligned} & \sum_{v \in V(\rho)} \text{vote}_v(M'(v), M(v)) \\ &= \sum_{(u,v) \in E(\rho) \cap M'} \text{vote}_u(v, M(u)) + \text{vote}_v(u, M(v)) \end{aligned}$$

Wegen (i) und der Betrachtung von G_M existieren keine mit $(1, 1)$ oder $(-1, -1)$ beschrifteten Kanten in ρ , daher sind alle Kanten mit $(1, -1)$ oder $(-1, 1)$ beschriftet und die Kantensumme, und damit die Summe der Voten der Knoten in ρ , ist 0.

- (b) Die Komponente ist ein alternierender Pfad, dessen zwei Endknoten jeweils von M' gepaart werden. Dann treffen die Betrachtungen von (a) unter Zuhilfenahme von Bedingung (ii) entsprechend zu und die Summe der Stimmen der Knoten in der Komponente ist 0.
- (c) Die Komponente ist ein alternierender Pfad, bei dem genau ein Endknoten von M' gepaart wurde. Dieser Fall entspricht dem um eine Kante in M erweiterten Fall (b). Diese zusätzliche Kante wird von dem Endknoten mit -1 bewertet (er prafert die Paarung durch M), sodass die Summe der Stimmen in der Komponente -1 ist.
- (d) Die Komponente ist ein alternierender Pfad, bei dem beide Endknoten von M gepaart wurden. Dieser Fall entspricht dem um zwei Kanten in M erweiterten Fall (b) unter Zuhilfenahme von Bedingung (iii). Diese Bedingung lasst maximal eine $(1, 1)$ -Kante zu, sodass

$$\sum_{(u,v) \in E(\rho) \cap M'} \text{vote}_u(v, M(u)) + \text{vote}_v(u, M(v)) \leq 2$$

ist. Die beiden von M gepaarten Endknoten prafieren ihre Paarung in M und fugen der Summe der Stimmen jeweils ein Stimmum von -1 hinzu, sodass fur die Gesamtsumme der Komponente gilt:

$$\begin{aligned} & \sum_{v \in V(\rho)} \text{vote}_v(M'(v), M(v)) \\ &= -2 + \sum_{(u,v) \in E(\rho) \cap M'} \text{vote}_u(v, M(u)) + \text{vote}_v(u, M(v)) \\ &\leq 0 \end{aligned}$$

Somit ist die Summe der Stimmen fur jede Zusammenhangskomponente von $M \oplus M'$ hochstens 0, woraus folgt, dass $\Delta(M', M) \leq 0$, was zu zeigen war.

Hinrichtung. Der zweite Teil des Beweises zeigt, dass ein Matching M , fur das eine der Bedingungen (i) bis (iii) nicht gilt, nicht popular ist. Hierzu wird fur jede der Bedingungen ein Matching M' konstruiert mit $M' \succ M$.

- (a) Gelte Bedingung (i) fur M nicht, dann existiert bezuglich M ein alternierender Kreis in G_M , der mindestens eine $(1, 1)$ -Kante enthalt. Sei die Menge der Kanten des Kreises C , so definiere $M' = M \oplus C$. Da C aufgrund der Betrachtung von G_M zwar keine $(-1, -1)$ -Kanten enthalt, jedoch mindestens eine $(1, 1)$ -Kante, und aufgrund der Alternanz im Kreis $|M| = |M'|$ gelten muss, ist die Summe der Stimmen der Knoten entlang C zugunsten M' groer 0, und somit $M' \succ M$.
- (b) Gelte Bedingung (ii) fur M nicht, dann existiert bezuglich M ein alternierender Pfad in G_M , der von einem in M isolierten Knoten ausgeht und mindestens eine $(1, 1)$ -Kante enthalt. Sei die Menge der

Kanten des Pfades P , so definiere $M' = M \oplus P$. Der Beweis erfolgt wie in (a), jedoch folgt hier $|M| \leq |M'|$ aus der Alternanz des Pfades sowie der Tatsache, dass mindestens einer der Anfangsknoten in M isoliert ist (vgl. Abbildung 2). Es folgt somit $M' \succ M$.

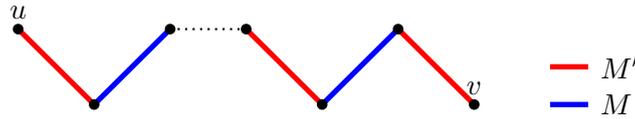


Abbildung 2: Alternierender Pfad mit bzgl. M isolierten Knoten u und v . Der Knoten v ist für diesen Beweis optional.

- (c) Gelte Bedingung (iii) für M nicht, dann existiert bezüglich M ein alternierender Pfad in G_M , der mindestens zwei $(1, 1)$ -Kanten enthält. Sei die Menge der Kanten des Pfades P , so ist $M' = M \oplus P$. Da beide Endpunkte von P von M gepaart werden, ist $|M| = |M'| + 1$. Da M jedoch zwei $(1, 1)$ -Kanten enthält und keine $(-1, -1)$ -Kante, ist erneut $M' \succ M$.

Es folgt, dass (i) bis (iii) hinreichende und notwendige Bedingung für die Popularität eines Matchings sind. \square

Dabei erlaubt es Bedingung (iii), dass ein populäres Matching – im Gegensatz zu stabilen Matchings – auch blockierende Kanten enthalten darf, die dann zur größeren Kardinalität beitragen können. Es fällt auf, dass Satz 1 auch für nicht-bipartite Graphen mit total geordneten Präferenzlisten gilt. Die Autoren stellen jedoch heraus, dass sich für solche Graphen nicht immer ein populäres Matching finden lässt.

Es folgt die hinreichende Bedingung für populäre Matchings maximaler Kardinalität, welche später Grundlage des Algorithmus bilden wird.

Satz 2. Sei G ein Graph und M ein populäres Matching auf G . Seien die Kanten $e \in E \setminus M$ von G mittels $\text{label}_M(e)$ beschriftet. Gibt es keinen augmentierenden Pfad bezüglich M in G_M , so ist M ein populäres Matching maximaler Kardinalität.

Beweis. Sei Q ein Matching mit $|Q| > |M|$. Zu zeigen ist, dass Q nicht populär ist.

Betrachte $Q \oplus M$. Da ein Matching Maximalgrad 1 hat, enthält $Q \oplus M$ ausschließlich (alternierende) Pfade und Kreise. Ein Kreis in $Q \oplus M$ enthält ebenso viele Kanten von Q wie von M ; wegen $|Q| > |M|$ muss daher ein M augmentierender Pfad in $Q \oplus M$ vorhanden sein, sei dies p .

Aufgrund des Satzes existiert in G_M jedoch kein augmentierender Pfad bezüglich M , also muss p wenigstens eine Kante enthalten, die mit $(-1, -1)$ beschriftet ist. Seien p_1, \dots, p_t diejenigen Pfade, die aus p durch Löschen

aller $(-1, -1)$ -Kanten entstehen. Beobachte, dass p_1 und p_t unter Satz 1 Bedingung (ii) fallen, also keine mit $(1, 1)$ beschriftete Kanten enthalten, während auf die restlichen $t - 2$ Pfade Satz 1 Bedingung (iii) zutrifft – sie enthalten jeweils höchstens eine mit $(1, 1)$ beschriftete Kante. Alle anderen Kanten sind mit $(-1, 1)$ bzw. $(1, -1)$ beschriftet.

Da p augmentierend ist, kann die Summe der Voten entlang p anhand der beschrifteten Kanten wie folgt zusammengefasst werden:

$$\sum_{v \in V(p)} \text{vote}_v(Q(v), M(v)) \leq \overbrace{2(t-2)}^{(1,1)\text{-Kanten}} - \overbrace{2(t-1)}^{(-1,-1)\text{-Kanten}} = -2$$

Es folgt, dass $Q \oplus p$ populärer ist als Q , was zu zeigen war. \square

4 Lösung

Der vorliegende Abschnitt stellt die bei [HK11] präsentierten Lösungen vor, einerseits zur Berechnung populärer Matchings maximaler Kardinalität, andererseits zum Test eines Matchings auf seine Popularität. Der Schwerpunkt soll hierbei in Hinblick auf den Umfang dieses Seminarbeitrags auf der ersten Lösung liegen. Der Algorithmus zum Test von Matchings auf Popularität wird zwar vorgestellt, jedoch wird auf Beweise zu Korrektheit und Laufzeit verzichtet.

4.1 Berechnung populärer Matchings maximaler Kardinalität

Zur Berechnung der gesuchten Matchings fanden die Autoren des Papers Kriterien, welche die hinreichende Bedingung in Satz 2 implizieren. Dieser Ansatz wird nachfolgend eingeführt; anschließend wird der resultierende Algorithmus beschrieben und Beweise dazu aufgezeigt.

4.1.1 Ansatz

Die Idee ist eine geschickte Neupartitionierung des Graphen G in $V = L \dot{\cup} R$. Erfolgt die Wahl der Partitionen richtig, lässt sich das gesuchte Matching dann mittels des Gale/Shapley-Algorithmus aus [GS62] bestimmen.

Definition 8. Sei $M \subseteq L \times R$ ein Matching; seien weiterhin alle Kanten $e \in E \setminus M$ mit $\text{label}_M(e)$ beschriftet. M heißt „günstig bezüglich (L, R) “, wenn es folgende zwei Bedingungen erfüllt:

- (i) Es existiert in $L \times R$ keine mit $(1, 1)$ beschriftete Kante.
- (ii) Alle Kanten in $L \times L$ sind mit $(-1, -1)$ beschriftet.

Die Eigenschaft (i) wird dabei von jedem stabilen Matching auf $L \times R$ erfüllt; ein geeignetes Matching lässt sich daher mit dem vorhandenen Algorithmus bestimmen. Problematisch ist korrekte Wahl von L , um Eigenschaft (ii) erfüllen zu können.

Satz 3. *Sei M ein Matching auf G . Ist M günstig bezüglich einer Partitionierung (L, R) von G , und ist M R -perfekt, so ist M ein populäres Matching maximaler Kardinalität.*

Beweis. Es ist also $M \subseteq L \times R$, die Kanten von G sind beschriftet mit label_M , und M ist günstig bezüglich (L, R) sowie R -perfekt. Zu zeigen: M erfüllt die Forderungen (i), (ii) und (iii) aus Satz 1 sowie die Forderung aus Satz 2. Diese vier Aspekte werden im folgenden getrennt gezeigt.

- (a) Sei C ein alternierender Kreis bezüglich M in G_M . Es ist zu zeigen, dass C keine $(1, 1)$ -Kante enthält. Da jede zweite Kante von C in $M \subseteq L \times R$ liegt, ist die Anzahl der L -Knoten und der R -Knoten in C gleich. Da G_M wegen der Günstigkeit von M jedoch keine Kanten aus $L \times L$ enthält, und C in G_M liegt, muss $C \subseteq L \times R$ sein. Daraus folgt wiederum mit der Günstigkeit von M die Behauptung.
- (b) Sei $p = (u_0, \dots, u_k)$ ein alternierender Pfad bezüglich M in G_M , dabei sei u_0 isoliert in M . Es ist zu zeigen, dass p keine $(1, 1)$ -Kante enthält. Da M R -perfekt ist, folgt $u_0 \in L$. Wegen $(L \times L) \cap E(G_M) = \emptyset$ folgt $u_1 \in R$. Wegen $M \subseteq L \times R$ folgt $u_2 \in L$. Wegen $(L \times L) \cap E(G_M) = \emptyset$ folgt erneut $u_3 \in R$, und so fort. p muss also in $L \times R$ liegen, und es folgt mit der Günstigkeit von M die Behauptung.
- (c) Sei $p = (u_0, \dots, u_k)$ ein alternierender Pfad bezüglich M in G_M . Es ist zu zeigen, dass p höchstens eine $(1, 1)$ -Kante enthält. Da Kanten in M gar nicht beschriftet werden, kann o.B.d.A. angenommen werden, dass $(u_0, u_1) \notin M$. Auch sei o.B.d.A. $u_0 \in R$, da der Beweis sonst wie in (b) wäre. Betrachte Knoten mit ungeradem Index. Wenn $\nexists i \in \mathbb{N}_0 : u_{2i+1} \in R$, wären in p keine Kanten aus $R \times R$ enthalten, es folgte $p \in L \times R$, und damit – wie oben – die Behauptung. Sei also nachfolgend i die kleinste Zahl, für die $u_{2i+1} \in R$ gilt (vgl. Abbildung 3). Damit ist $(u_{2i}, u_{2i+1}) \in R \times R$. Wegen der Alternanz des Pfades ist dann $(u_{2i+1}, u_{2i+2}) \in M \subseteq L \times R$, also $u_{2i+2} \in L$. Da G_M keine Kanten aus $L \times L$ enthält, ist aber wieder $(u_{2i+2}, u_{2i+3}) \in L \times R$, und $(u_{2i+3}, u_{2i+4}) \in M \subseteq L \times R$, und so fort. Damit ist die einzige möglicherweise mit $(1, 1)$ beschriftete Kante (u_{2i}, u_{2i+1}) , und es folgt die Behauptung.
- (d) Sei $p = (u_0, \dots, u_k)$ ein augmentierender Pfad bezüglich M in G_M , es sind also u_0, u_k isoliert in M . Beobachte, dass bei einem alternierenden, augmentierenden Pfad jede Kante, die nicht in M liegt, in einem Knoten

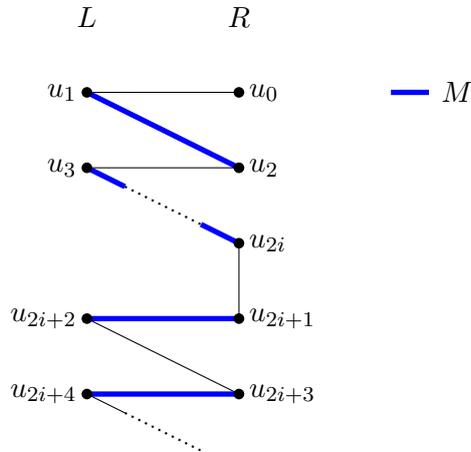


Abbildung 3: Alternierender Pfad des bezüglich (L, R) günstigen sowie R -perfekten Matchings M .

mit ungeradem Index endet, k also ungerade sein muss. Beobachte weiterhin gemäß der Argumentation aus (b), dass jeder Knoten mit geradem Index in L , und jeder Knoten mit ungeradem Index in R liegen muss. Also liegt $u_k \in R$, ein Widerspruch zur R -Perfektheit von M . Es folgt die Behauptung.

Damit ist M ein populäres Matching maximaler Kardinalität, was zu zeigen war. \square

Der Satz 3 liefert damit sehr konkrete Kriterien für den im folgenden Abschnitt gezeigten Algorithmus, um ein populäres Matching maximaler Kardinalität zu berechnen.

4.1.2 Algorithmus

Die Aufgabe des Algorithmus ist es nun, zu einem gegebenen bipartiten Graphen $G = (\mathcal{A} \cup \mathcal{B}, E)$ eine neue Partitionierung (L, R) sowie ein Matching M zu finden, so, dass M günstig bezüglich dieser Partitionierung sowie R -perfekt ist.

Wie bereits erläutert, soll hierbei der Gale/Shapley-Algorithmus genutzt werden, um Eigenschaft (i) eines günstigen Matchings herzustellen. Noch zu beantworten ist aber die Frage, wie Eigenschaft (ii) sowie die R -Perfektheit zu erreichen sind.

Die Idee hierzu: der Algorithmus konstruiert die Partition schrittweise. Er beginnt dabei im ersten Schritt damit, der linken Knotenmenge L_1 genau jene Knoten zuzuweisen, die in allen stabilen Matchings ungepaart bleiben². Es ist

²Wie erläutert, lassen alle stabilen Matchings zu G die selbe Knotenmenge ungepaart.

zu beobachten, dass die Knotenmenge L_1 unabhängig ist. Die sich ergebende Knotenmenge $R_1 = V \setminus L_1$ muss nun schrittweise so reduziert werden, dass der Gale/Shapley-Algorithmus ein R -perfektes Matching findet; es müssen also solche Knoten, die nicht gepaart werden, nach L verschoben werden. Um nun Eigenschaft (ii) zu wahren, werden dabei von Schritt zu Schritt stets nur Frauen verschoben. Dies hat zur Folge, dass sich der Fortschritt des Algorithmus durch die Zahl der nach links verschobenen Frauen bestimmen lässt. Die Möglichkeit einer Lösung, in der Männer auf der linken Seite vorkommen, wird lediglich in Zwischenschritten geprüft.

Der Vorgang ist in Algorithmus 1 dargestellt. Es folgt eine knappe Beschreibung dieses Algorithmus, der Nachweis seiner Terminierungseigenschaften, Angaben zur Laufzeit sowie ein Korrektheitsbeweis.

Eingabe : $G = (V = \mathcal{A} \cup \mathcal{B}, E)$ mit tot. geordn. Präferenzlisten

Ausgabe : Populäres Matching maximaler Kardinalität auf G

```

1  $S \leftarrow \text{GALE/SHAPLEY}(\mathcal{A}, \mathcal{B})$ 
2  $L_1 \leftarrow \{v \in V \mid S(v) = \varepsilon\}$ ;  $R_1 \leftarrow V \setminus L_1$ 
3  $i \leftarrow 1$ 
4 forever do
5    $M_i \leftarrow \text{GALE/SHAPLEY}(L_i, R_i)$ 
6   if  $M_i$  ist  $R_i$ -perfekt then
7     | return  $M_i$ 
8   end
9    $A_i \leftarrow \{a \in \mathcal{A} \cap R_i \mid M_i(a) = \varepsilon\}$ 
10   $L'_i \leftarrow L_i \cup A_i$ ;  $R'_i \leftarrow V \setminus L'_i$ 
11   $M'_i \leftarrow \text{GALE/SHAPLEY}(L'_i, R'_i)$ 
12  if  $M'_i$  ist  $R_i$ -perfekt then
13    | return  $M'_i$ 
14  end
15   $B_i \leftarrow \{b \in R'_i \mid M'_i(b) = \varepsilon\}$ 
16   $L_{i+1} \leftarrow L_i \cup B_i$ ;  $R_{i+1} \leftarrow V \setminus L_{i+1}$ 
17   $i \leftarrow i + 1$ 
18 end

```

Algorithmus 1: Berechnung eines populären Matchings maximaler Kardinalität

Beschreibung. Der Algorithmus beginnt mit Berechnung einer initialen Partitionierung (L_1, R_1) in Zeile 1, indem ein stabiles Matching S berechnet und nun die beiden Partitionen aus darin ungepaarten bzw. gepaarten Knoten gebildet werden. Die Variable i wird als Rundenzähler genutzt.

Es folgt die Schleife ab Zeile 4, welche jeweils das Matching für die aktuelle Runde berechnet, dann – sollte dies noch nicht das gesuchte Matching sein

– ein Hilfsmatching bestimmt, und anschließend – sollte auch dieses Hilfsmatching nicht alle geforderten Eigenschaften erfüllen – die Partitionierung für den nächsten Schritt ermittelt.

Dem Hilfsmatching M'_i (Zeile 9ff.) liegt dabei die Partitionierung (L'_i, R'_i) zugrunde, in welcher alle Männer aus R_i , die in M_i isoliert sind, der linken Seite L'_i zugeschlagen wurden.

Anhand dieses Hilfsmatchings wird dann ab Zeile 15 die Menge der Frauen bestimmt, welche sich in der folgenden Runde $i + 1$ auf der linken Seite befinden sollen – es sind wiederum diejenigen Frauen in R'_i , welche in M'_i isoliert sind. Es wird dabei noch gezeigt, dass B_i tatsächlich nur Frauen enthält.

Stellt sich das Matching M_i als R_i -perfekt heraus, so ist (L_i, R_i) die gesuchte Partitionierung und das Matching wird zurückgegeben. Ist ein M'_i R'_i -perfekt, so ist die gesuchte Partitionierung entsprechend (L'_i, R'_i) und die Rückgabe erfolgt mit diesem Matching.

Terminierung. Der Algorithmus läuft rundenbasiert ab, dabei wird in jeder Runde ein Fortschritt zur Lösung hin gemacht. Wie oben erwähnt, erfolgt der Fortschritt über das Verschieben von Frauen auf die linke Seite. Zuerst soll hierzu nachgewiesen werden, dass es tatsächlich Frauen sind, welche in Zeile 15 nach links verschoben werden.

Hilfssatz 1. *Es gilt für alle $i \geq 1$, dass $B_i \subseteq \mathcal{B}$.*

Beweis. Es ist B_i die Menge derjenigen Knoten, welche in M'_i isoliert sind; die zugehörige Partitionierung (L'_i, R'_i) entsteht aus (L_i, R_i) , wenn alle in M_i isolierten Männer (A_i) auf die linke Seite verschoben werden.

Beobachte, dass jeder Knoten aus R_i , der in M_i gepaart ist, auch in M'_i einen Partner findet: das Entfernen ungepaarter Männer von der rechten Seite ändert hieran nichts, und nach dem Hinzufügen dieser Männer zur linken Seite finden Knoten auf der rechten Seite wenigstens ihren Partner aus M_i , wenn nicht sogar einen stärker präferierten Partner aus A_i .

Da R'_i keine in M_i ungepaarten Männer mehr enthält, und alle Knoten, die in R_i bezüglich M_i gepaart waren, auch in R'_i bezüglich M'_i gepaart sind, müssen alle Knoten in B_i Frauen sein. \square

Mit diesem Hilfssatz kann nun eine obere Schranke für die Zahl der Runden angegeben werden.

Hilfssatz 2. *Die Zahl der Durchläufe für die Schleife in Algorithmus 1, Zeile 4 ist höchstens $|\mathcal{B}|$.*

Beweis. Beobachte: $B_i \neq \emptyset$, da sonst M'_i R'_i -perfekt wäre. Es folgt mit Hilfssatz 1, dass in jeder Runde mindestens eine Frau von rechts nach links bewegt wird. Da niemals Frauen von links nach rechts bewegt werden, ist die Zahl der Runden und damit der Schleifendurchläufe auf $|\mathcal{B}|$ begrenzt. \square

Laufzeit. Der Aufruf GALE/SHAPLEY betrachtet jede Kante einmal und läuft in $\mathcal{O}(m)$. Es kann angenommen werden, dass die Prüfung auf R -Perfekteit, Mengenoperationen sowie das Bilden der Mengen ungepaarter Knoten höchstens in $\mathcal{O}(n)$ ablaufen. Daraus lässt sich die Laufzeit des Algorithmus wie folgt bestimmen:

Hilfssatz 3. Die Laufzeit des Algorithmus 1 liegt in $\mathcal{O}(m \cdot n_0)$, für $n_0 = \min(|\mathcal{A}|, |\mathcal{B}|)$.

Beweis. Während eines Schleifendurchlaufs erfolgen seriell zwei Aufrufe von GALE/SHAPLEY, zwei Prüfungen auf R -Perfekteit, zwei Bestimmungen ungepaarter Knoten sowie vier Vereinigungen / Differenzen auf Mengen. Die Laufzeit pro Schleifendurchlauf beträgt daher $\mathcal{O}(n + m)$ und mit $m \geq \frac{n}{2}$ auch $\mathcal{O}(m)$.

Mit Hilfssatz 2 und der Annahme, dass $|\mathcal{B}| < |\mathcal{A}|$, bzw. der Definition von n_0 folgt, dass die Gesamtlaufzeit des Algorithmus in $\mathcal{O}(m \cdot n_0)$ liegt. \square

Korrektheit. Zum Beweis der Korrektheit des Algorithmus ist zu zeigen, dass das zurückgegebene Matching M günstig bezüglich einer Partition (L, R) des Graphen, sowie R -perfekt ist. Letzteres ist trivial; es verbleibt der Beweis der Günstigkeit von M_i und M'_i . Dabei wird Eigenschaft (i) von Definition 8 konstruktiv aufgrund der Funktionsweise des GALE/SHAPLEY-Algorithmus erfüllt. Eigenschaft (ii) wird induktiv in vier Hilfssätzen gezeigt.

Hilfssatz 4. $\forall e \in E \cap (L_1 \times L_1) : \text{label}_{M_1}(e) = (-1, -1)$.

Beweis. L_1 besteht aus der Menge der in jedem stabilen Matching ungepaarten Männer A_0 sowie der in jedem stabilen Matching ungepaarten Frauen B_0 . Beobachte, dass L_1 folglich unabhängig sein muss, wodurch die geforderte Eigenschaft erfüllt ist. \square

Hilfssatz 5. $\forall e \in E \cap (L'_1 \times L'_1) : \text{label}_{M'_1}(e) = (-1, -1)$.

Beweis. Es ist $L'_1 = A_0 \cup A_1 \cup B_0$. Sei $e = (a, b)$ und o.B.d.A. $a \in \mathcal{A}$ und $b \in \mathcal{B}$. Es folgt $b \in B_0$, und da keine Kanten zwischen B_0 und A_0 existieren, sei nachfolgend $a \in A_1$.

A_1 ist jedoch die Menge derjenigen Männer aus R_1 , die in M_1 isoliert sind. a wird folglich nicht von b gegenüber $M'_1(b)$ bevorzugt, und es ist $\text{vote}_b(a, M'_1(b)) = -1$.

b ist in jedem stabilen Matching ungepaart und wird folglich nicht von a gegenüber den Knoten in R'_1 bevorzugt, es ist daher auch $\text{vote}_a(b, M'_1(a)) = -1$. \square

Mit Induktionshypothese $\forall e \in E \cap (L'_{i-1} \times L'_{i-1}) : \text{label}_{M'_{i-1}}(e) = (-1, -1)$ für $i \geq 2$ erfolgt nun der Induktionsschritt.

Hilfssatz 6. $\forall e \in E \cap (L_i \times L_i) : \text{label}_{M_i}(e) = (-1, -1)$.

Beweis. Es ist $L_i = A_0 \cup B_0 \cup \dots \cup B_{i-1}$. Sei $e = (a, b)$ und o.B.d.A. $a \in \mathcal{A}$ und $b \in \mathcal{B}$.

Betrachte die Präferenzen von a . Beobachte, dass aufgrund der Induktionshypothese zwischen A_0 und $B_0 \cup \dots \cup B_{i-2}$ alle Kanten bezüglich M'_{i-1} mit $(-1, -1)$ beschriftet sind. B_{i-1} ist die Menge derjenigen Frauen, die von $A_0 \cup A_{i-1}$ nicht gegenüber anderen Knoten in R'_{i-1} bevorzugt wurde; das Rückverschieben von A_{i-1} auf die rechte Seite schränkt dabei die Auswahl der Partner für A_0 nicht ein. Es folgt, dass $\text{vote}_a(b, M_i(a)) = -1$.

Betrachte nun die Präferenzen von b . a ist ein Knoten, der in keinem stabilen Matching gepaart wird, sodass b Knoten aus $\mathcal{A} \setminus A_0$ bevorzugt. Es folgt, dass $\text{vote}_b(a, M_i(b)) = -1$. \square

Bevor die Günstigkeit von M'_i gezeigt werden kann, wird festgestellt, dass $A_i \subseteq A_{i-1}$.

Beweis. Betrachte die Situation in (L_i, R_i) . Gegenüber der Partitionierung im vorhergehenden Schritt, (L_{i-1}, R_{i-1}) wurde B_{i-1} auf die linke Seite verschoben. Beobachte, dass hierdurch jeder Mann in R_i mindestens von so vielen Frauen in L_i gewählt wird, wie in L_{i-1} ; dass also gilt: ein Mann, der in M_{i-1} gepaart war, bleibt dies in M_i .

Es folgt $A_i \subseteq A_{i-1}$. \square

Hilfssatz 7. $\forall e \in E \cap (L'_i \times L'_i) : \text{label}_{M'_i}(e) = (-1, -1)$.

Beweis. Es ist $L'_i = A_0 \cup A_i \cup B_0 \cup \dots \cup B_{i-1}$. Sei $e = (a, b)$ und o.B.d.A. $a \in \mathcal{A}$ und $b \in \mathcal{B}$.

Betrachte die Präferenzen von b . a liegt entweder in A_0 oder in A_i . Da A_0 die Menge derjenigen Knoten ist, die in keinem stabilen Matching gepaart werden, und somit nicht von b bevorzugt werden, sei nachfolgend a in A_i . A_i ist jedoch die Menge derjenigen Knoten aus R_i , welche in M_i ungepaart blieben, sodass $\text{vote}_b(a, M_i(b)) = -1$. Das Verschieben der Knoten in A_i ändert an der Auswahl der Partner für b nichts, sodass $\text{vote}_b(a, M'_i(b)) = -1$.

Betrachte nun die Präferenzen von a . Da B_0 die Menge derjenigen Knoten ist, die in keinem stabilen Matching gepaart werden, und somit nicht von a bevorzugt werden, sei nachfolgend b in $B_1 \cup \dots \cup B_{i-1}$. Es gilt für $j \geq 1$, dass B_j von keinem Knoten in A_j gegenüber anderen Knoten in R'_j bevorzugt wurde; da $A_i \subseteq A_{i-1}$ gilt also auch, dass a auf der rechten Seite bessere Partner fand. Die Verkleinerung der Menge der Männer auf der linken Seite verschlechtert dabei die Auswahl für a nicht. Es folgt, dass $\text{vote}_a(b, M'_i(a)) = -1$. \square

Dies zeigt, dass für alle Runden des Algorithmus $i \geq 1$ die Matchings M_i und M'_i alle Eigenschaften der Günstigkeit bzgl. der jeweiligen Partitionierung erfüllen. Hilfssatz 2 besagt, dass ein solches Matching, das zudem R -perfekt ist, nach höchstens $|\mathcal{B}|$ Runden gefunden wird. Damit, sowie mit Satz 3, folgt der zentrale Satz in [HK11]:

Satz 4. Zu einem gegebenen bipartiten Graphen $G = (\mathcal{A} \cup \mathcal{B}, E)$ mit total geordneten Präferenzlisten kann ein populäres Matching maximaler Kardinalität in $\mathcal{O}(m \cdot n_0)$ berechnet werden, für $n_0 = \min(|\mathcal{A}|, |\mathcal{B}|)$.

4.2 Test eines Matchings auf Popularität

Aus den Bedingungen in Satz 1 entwickeln die Autoren des betrachteten Papers einen Algorithmus, um für ein gegebenes Matching in Linearzeit zu entscheiden, ob dieses Matching populär ist oder nicht. Den Autoren war bis dahin lediglich ein Verfahren in $\mathcal{O}(m\sqrt{n})$ bekannt.

Da der Schwerpunkt dieses Seminarbeitrags auf dem in Unterabschnitt 4.1 dargestellten Algorithmus zur Berechnung populärer Matchings maximaler Kardinalität liegt, wird hier lediglich der Testalgorithmus selbst dargestellt. Auf Nachweise zur Korrektheit und Laufzeit wird verzichtet, diese sind allerdings auch vergleichsweise leicht ersichtlich.

Der Algorithmus soll zu einem gegebenen Graphen $G = (\mathcal{A} \cup \mathcal{B}, E)$ mit total geordnete Präferenzlisten und einem zugehörigen Matching M entscheiden, ob dieses populär ist oder nicht. Die drei Bedingungen des Satzes 1 lassen sich dabei durch Betrachten des Graphen und des Matchings entscheiden. Der Algorithmus von [HK11] geht hierzu geschickt vor, indem er, ausgehend von zu blockierenden Kanten inzidenten Knoten jeweils solche Kanten betrachtet, die abwechselnd zu M gehören bzw. nicht gehören. Dadurch verfolgt er alternierende Pfade; die drei Eigenschaften lassen sich dann leicht entscheiden. Die lineare Laufzeit folgt, da pro Graphdurchlauf jeder Knoten nur einmal betrachtet werden muss.

Das Vorgehen ist wie folgt.

1. Markiere die Endpunkte aller Kanten in G_M , die bezüglich M mit $(1, 1)$ beschriftet wurden.
2. Konstruiere einen Baum; verteile die Knoten von G_M wie folgt auf Ebenen:
 - Platzierte markierte Frauen auf Ebene 0.
 - Dann, für $k \geq 0, i = 2k + 1$:
 - Platziere solche Männer auf Ebene i , welche mit Knoten in Ebene $i - 1$ gepaart sind.
 - Platziere solche Frauen auf Ebene $i + 1$, welche mit Knoten in Ebene i benachbart sind und bisher nicht betrachtet wurden.
 - Der Vorgang endet, wenn keine Knoten mehr zu betrachten sind.
3. Wurde ein markierter Mann oder eine ungepaarte Frau angetroffen, gebe „unpopulär“ aus und Ende.

4. Konstruiere erneut einen Baum; verteile die Knoten von G_M wie oben auf Ebenen, beginne dabei jedoch mit markierten *Männern* in Ebene 0, deren Partnern in Ebene 1 usw.
Wurde hierbei ein ungepaarter Mann angetroffen, gebe „*unpopulär*“ aus und Ende.
5. Gebe „*populär*“ aus.

Beachte, dass die hier konstruierten Bäume folgende Eigenschaft aufweisen: zu jedem in ihnen enthaltenen Knoten v existiert ein alternierender Pfad von einem Knoten in Ebene 0 zu v , und somit ein alternierender Pfad, welcher mit einer blockierenden Kante beginnt. Ein markierter Mann in Schritt 3 ist dabei Beweis für eine zweite blockierende Kante (und somit Satz 1, Bedingung (iii)) oder einen Kreis mit blockierender Kante (und somit Satz 1, Bedingung (i)). Ein ungepaarter Knoten hingegen ist jeweils Beweis für einen alternierenden Pfad, welcher von einem ungepaarten Knoten aus startet (und somit Satz 1, Bedingung (ii)).

5 Zusammenfassung

[HK11] betrachtet Matchings auf Graphen mit total geordneten Präferenzlisten und zeigt für beliebige solcher Graphen hinreichende und notwendige Bedingungen für die Popularität eines Matchings auf. Zudem beschreibt die Veröffentlichung eine hinreichende Bedingung zur maximalen Kardinalität populärer Matchings. Hiermit können möglichst große Matchings berechnet werden, welche von der Mehrheit der Knoten bevorzugt werden.

Die Veröffentlichung präsentiert einen effizienten Algorithmus zur Berechnung solcher Matchings für bipartite Graphen, sowie zum Test eines Matchings auf Popularität.

Literatur

- [GS62] GALE, David ; SHAPLEY, Lloyd S.: College Admissions and the Stability of Marriage. In: *The American Mathematical Monthly* Bd. Vol. 69, No. 1, 1962, S. 9–15
- [HK11] HUANG, Chien-Chung ; KAVITHA, Telikepalli: Popular Matchings in the Stable Marriage Problem. In: *Proceedings of the ICALP 2011*, 2011, S. 666–677