comput. complejo.15 (2006), 20 - 39
1016-3328 / 06 / 010020–20
DOI 10.1007 / s00037-006-0205-6

complejidad computacional

SOBRE LA COMPLEJIDAD DE APROXIMAR k-CONFIGURAR EMBALAJE

Elad Hazan, Shmuel Safra y Oded Schwartz

Resumen. Dado un k-hipergrafo uniforme, el Máximo k-Conjunto de embalaje El problema es encontrar el máximo conjunto de bordes disjuntos. Demostramos que este problema no puede aproximarse eficientemente dentro de un factor de Ω (k / en k) a menos que P = NP. √ThEsto mejora la dureza anterior del factor de aproximación de k /2O (en k) por Trevisan. Este resultado se extiende a el problema de k-Adaptación dimensional.

Palabras clave. Complejidad computacional, dureza de aproximación, empaquetamiento de conjuntos.

Clasi fi cación de materias. 68Q17.

1. Introducción

Este artículo estudia el siguiente problema básico de optimización: dada una familia de conjuntos sobre un cierto dominio, encuentre el número máximo de conjuntos disjuntos. Consideramos el caso en el que todos los conjuntos de la familia dada son del mismo tamañok.

Para el caso donde k = 2, podemos ver los conjuntos como aristas en un gráfico cuyo los vértices son el dominio y, por lo tanto, el problema es exactamente el famoso problema de coincidencia máxima que se puede resolver en tiempo polinomial (Papadimitriou 1994). Parak ≥ 3, nuevamente viendo los conjuntos como hiper-bordes en un hiper-gráfico, el problema de encontrar la máxima coincidencia en k-Los hipergráficos uniformes son NP-hard. Por lo tanto, a menos que P = NP, la mejor esperanza es obtener un algoritmo de aproximación de tiempo polinomial con garantía de aproximación probadamente buena.

El algoritmo codicioso simple es el siguiente: elija iterativamente un conjunto arbitrario y agréguelo a la colección de conjuntos mantenidos hasta ahora, mientras elimina todos los conjuntos que lo cruzan. Continúe mientras queden bordes en el gráfico. Obviamente, este algoritmo devuelve una familia de conjuntos disjuntos por pares. Es fácil demostrar que este algoritmo proporciona unak-aproximación a la solución óptima. Una mejora constante en la relación de aproximación, ak /2 (Hurkens y Schrijver 1989), se puede obtener mediante una simple heurística de búsqueda local y es la aproximación más conocida hasta la fecha.

En este trabajo demostramos que esta última garantía de aproximación es casi ajustada, demostrando lo siguiente:

Teorema 1.1. Es NP-difícil de aproximar k-SP a dentro Ω (k /en k).

1.1. Resultados anteriores.El general Embalaje de conjunto máximo el problema es de la siguiente manera: dada una familia F = {$S_1, \ldots, S_{metro}$} de conjuntos en un determinado dominio D = {$X_1, \ldots, X_{norte}$}, el objetivo es encontrar un empaquetamiento máximo, es decir, un número máximo de conjuntos disjuntos por pares de la familia dada. Este problema es a menudo Expresado en la terminología de la teoría de grafos, como un sistema de conjuntos es de hecho un hipergrafo donde los vértices son los elementos del dominio y las aristas son los conjuntos dados. En la jerga de la teoría de grafos, un conjunto disjunto de aristas se denomina coincidencia, por lo que el objetivo es encontrar una coincidencia máxima.

Los problemas de empaque se encuentran entre los problemas fundamentales de optimización combinatoria. Variantes deEmbalaje del conjunto máximo, incluyendo el Conjunto independiente máximo y Máxima camarilla problemas, han sido ampliamente estudiados (Arora et al. 1998; Arora y Safra 1998; Bar-Yehuda y Moran 1984; Boppana y Halld´órsson 1992; Feigeet al. 1996; H˚ åstad 1999; Wigderson 1983). Estas formulaciones generales de problemas de empaque son notoriamente difíciles incluso de aproximar: H˚åstad (1999) demostró que Máxima camarilla (y ahí- delantero Conjunto independiente máximo y Embalaje de conjunto máximo también) no se puede aproximar a dentro EN$^{1-ε}$) a menos que NP ⊆ ZPP (para cada ε> 0). El mejor algoritmo de aproximación paraConjunto independiente máximo logra una relación de aproximación de EN/Iniciar sesión^2 N) (Boppana y Halld´órsson 1992).

En este artículo consideramos varias variantes naturales de los problemas de empaque. El primero, y quizás el más natural, es cuando el tamaño de los hiperbordes está delimitado pork. Este problema se llama Máximo k-Establecer embalaje (para abreviar k-SP). Si además delimitamos el grado de los vértices por dos, esto se convierte en el problema del máximo conjunto independiente en gráficas de grado como máximo.k.

Otra restricción natural (más fuerte) estudiada aquí es cuando imponemos un límite a la colorabilidad del gráfico de entrada. Este es el problema deMáximo k-Coincidencia dimensional (para abreviar k-DM). Es una variante deMáximo k-Conjunto de embalaje donde los vértices del hipergráfico de entrada son una unión de k conjuntos disjuntos, V = $V_1 \cup \cdots \cup V_k$, y cada hiperfilo contiene exactamente un vértice de cada conjunto, es decir, mi ⊆ $V_1 \times \cdots \times V_k$. En otras palabras, los vértices del hipergráfico se pueden colorear usando k colores, de modo que ningún hiperfilo contenga el mismo color dos veces. Un gráfico que tiene esta propiedad se llamak-fuertemente colorante. Así, la versión delimitada por colores de Máximo k-Conjunto de embalaje es, dado un k-uniforme k-hipergráfico fuertemente colorante, encuentre una coincidencia de tamaño máximo.

Estas variantes limitadas de Embalaje de conjunto máximo se sabe que admiten algoritmos de aproximación mejor que sus versiones generales, la calidad de la la aproximación es una función de los límites. Como se mencionó anteriormente, el algoritmo codicioso garantiza unk-aproximación para Máximo k-Establecer embalaje. Una simple heurística de búsqueda local logra una relación de aproximación de k /2 (Hurkens y Schrijver 1989). Este es, hasta la fecha, el mejor algoritmo de aproximación ritmo para Máximo k-Coincidencia dimensional también.

Para el caso especial donde k = 2, ambos problemas se pueden resolver en polinomio hora. Coincidencia máxima bidimensional es solo el problema de encontrar una coincidencia máxima en un gráfico bipartito, y puede resolverse en tiempo polinomial, por ejemplo, mediante una reducción a problemas de flujo de red (Papadimitriou 1994). Embalaje máximo de 2 juegos Es el problema de encontrar una coincidencia máxima en un gráfico general, y para este problema también se conocen algoritmos de tiempo polinomial (Edmonds 1965) (para algoritmos eficientes recientes, ver Mucha y Sankowski 2004).

Sin embargo, para todos k ≥ 3, Máximo k-Coincidencia dimensional es NP-difícil (Karp 1972; Papadimitriou 1994). Además, parak = 3, se sabe que el problema es APX-hard (Kann 1991). soloet al. (1995) demostró que para un k, Máximo k-Conjunto independiente (encontrar un conjunto independiente de tamaño máximo en k-gráficos regulares, para abreviar k-IS) es NP-difícil de aproximar a dentro k^c para algunos c> 0. Esto fue más tarde √ soypagviajó a lo mejor actualmente factor de inaproximabilidad asintótico de k $/2^{O(\text{en } k)}$ (Trevisan 2001). Todo duro factores de ness para Máximo k-Conjunto independiente sostener de hecho para Máximo k + 1-Coincidencia dimensional también (mediante una simple reducción).

El algoritmo de aproximación más conocido para k-IS logra una relación de aproximación de OK registro de registro k /Iniciar sesión k) (Vishwanathan 1996). Parak-ES de baja k valores, el mejor algoritmo de aproximación logra una relación de aproximación de (k + 3) / 5 para k ≥ 3 (Berman y Fujito 1995; Berman y Furer 1994). Berman y Karpinski (2003) mostraron un factor de inaproximabilidad de 98/97 paraMáximo 3-Coincidencia dimensional. Para más información sobre los resultados de inapropiabilidad de bajo grado, consulte Hazan (2002).

1.2. Nuestro aporte.Mejoramos el factor de inaproximación para Máximo k-Conjunto de embalaje, y demostrar que es NP-difícil de aproximar k-SP dentro de Ω (k /en k). Extendemos este resultado a Máximo k-Coincidencia dimensional. Estos resultados también implican el mismo límite para (k + 1) -Gráficos sin garras (ver Halld´órsson 1998 para la definición de este problema y su relación con k-SP). Sin embargo, no se mantienen pork-ES.

La prueba de estos límites inferiores introduce un objeto combinatorio llamado hiper-borde-dispersor, y presentamos una construcción aleatoria de dicho objeto. Este objeto puede ser de interés independiente.

1.3. Esquema.En la Sección 2 se dan algunos preliminares. La Sección 2.2 presenta la noción de hiperdispersores de borde. La sección 3 contiene la prueba de la dureza asintótica de aproximación parak-SP. La sección 4 extiende la prueba para que sea válidak-DM. La existencia de un buen hiperdispersor se demuestra en la Sección 5. La optimización de sus parámetros se muestra en la Sección 5.1. La sección 6 contiene una discusión sobre las implicaciones de nuestros resultados, las técnicas utilizadas y algunos problemas abiertos.

2. Preliminares

Para probar la inapropiabilidad de un problema de maximización, generalmente se define un problema de brecha correspondiente.

Definición 2.1. Dejar A ser un problema de maximización. brecha-A- [a, b] es el siguiente problema de decisión: Dada una instancia de entrada, decida si

- existe una solución de tamaño fraccionario al menos B, o
- cada solución de la instancia dada es de tamaño fraccional menor que una.

Si el tamaño de la solución se encuentra entre estos valores, la salida no tiene restricciones.

Claramente, para cualquier problema de maximización, si la brechaA- [a, b] es NP-duro, que NP-difícil de aproximar A dentro de cualquier factor menor que licenciado en Letras.

Nuestro principal resultado en este artículo se deriva de una reducción del siguiente problema.

Definición 2.2. MAX-3-LIN-q es el siguiente problema de optimización: Entrada: Un conjunto Φ de ecuaciones lineales módulo un entero q, cada uno dependiendo de tres variables.
Problema: Encuentre una tarea que satisfaga el número máximo de ecuaciones.

El siguiente teorema central surge de una larga línea de investigación, utilizando el teorema PCP (Arora et al. 1998; Arora & Safra 1998) y el teorema de la repetición paralela (Raz 1998) como punto de partida:

Teorema 2.3 (H˚ åstad 2001). Para cada q ∈ norte y ε> 0, espacio-MAX-3-LIN-q- [1 /q + ε, 1 - ε] es NP-duro. Además, el resultado es válido para instancias de MAX-3-LIN-q en el que el número de ocurrencias de cada variable es una constante (dependiendo de ε y en q).

Denotamos una instancia de MAX-3-LIN-q por $\Phi = \{\phi_1, \ldots, \phi_{metro}\}$. Φ ha terminado el conjunto de variables $X = \{x_1, \ldots, X_{norte}\}$. Deje Φ (X) ser el (multi) conjunto de todas las ecuaciones en Φ dependiendo de X ∈ X (es decir, puede verse como todas las ocurrencias de X). Denotar por Sat (Φ, A) el conjunto de todas las ecuaciones en Φ satisfecho por la asignación UNA. Para una tarea A, denotamos por $A|_X$ el valor a ∈ [q] que A asigna a X.

2.1. Hipergráficos.Un hipergrafo H = (V, E) consta de un conjunto de vértices V y una colección mi de subconjuntos de V llamados hiper-bordes (para abreviar, bordes).

Como de costumbre, el grado de un vértice es el número de aristas en las que aparece. Un hipergráfico H se llama D-regular si el grado de cada uno de sus vértices es exactamente D, y k-uniforme si el tamaño de cada uno de sus bordes es exactamente k.

A pareo es un subconjunto METRO de mi tal que todos los bordes de METRO son disjuntos por pares. Usamos la siguiente definición no estándar de un conjunto independiente en hipergráficos:

Definición 2.4. Dejar H = (V, E) ser un hipergrafo. Un subconjunto de vérticesI ⊆ V se llama un conjunto independiente si alguna ventaja mi ∈ mi contiene como máximo un vértice de I.

De él derivamos la definición correspondiente (pero no estándar) de colorabilidad:

Definición 2.5. El hipergrafo H = (V, E) se ha dicho k-fuertemente colorante si hay una partición de V dentro k conjuntos de modo que cada parte sea un conjunto independiente.

Por lo tanto, un k-uniforme k-hipergráfico fuertemente colorante H puede ser denotado por $H = (V_1, \ldots, V_k, MI)$, dónde $mi \subseteq V_1 \times \cdots \times V_k$. Una noción análoga a la colorabilidad fuerte se aplica a los bordes de un hipergráfico:

Definición 2.6. Un hipergrafo H = (V, E) se ha dicho D-fuertemente coloreado si existe una coloración de los bordes f: E → [D] de modo que cada vértice participa como máximo en una arista de cada color.

Usando estas de fi niciones podemos definir formalmente el problema de empaquetamiento relacionado estudiado aquí:

Definición 2.7. Máximok-Conjunto de embalaje es la siguiente optimización problema:

Entrada: A k-hipergráfico uniforme $H = (V_1, \ldots, V_k, MI)$.

Problema: encuentre una coincidencia de tamaño máximo en H.

Máximo k-Coincidencia dimensional es el mismo problema, donde la el gráfico de entrada es k-fuertemente colorante.

2.2. Hiperdispersores. La siguiente definición es una generalización de dis- gráficos perser. Para de fi niciones y resultados con respecto a los dispersores, consulte Radhakrishnan y Ta-Shma (2000).

Definición 2.8. Un hipergrafo H = (V, E) es un (q, δ) -dispersor de hiper-borde Si existe una partición de sus bordes: $E = E_1 \cup \cdots \cup mi_q$ con $|mi_1| = \cdots = |mi_q|$, de tal manera que cada gran coincidencia METRO de H está (casi) concentrado en una parte de Los bordes. Formalmente, para cada METRO existe I así que eso

$$|ME_{yo}| \leq \delta \, |E|.$$

Lema 2.9. Para cada q> 1 y t> c (q) (dónde c (q) es una constante que depende solo de q) existe un hipergrafo H = (V, E) tal que

- V = [t] × [D], dónde d = Θ (q en q).
- H es un (q, 1 /q^2) -dispersor de hiper-borde.
- H es D-uniforme, D-fuertemente colorante.
- H es q-regular, q-fuertemente colorante en los bordes.

De ahora en adelante denotamos tal grafo por D[t, q]. Debido a las condiciones de regularidad, uniformidad y colorabilidad de este hipergráfico, el número de aristas es exactamente qt, y se pueden dividir en q conjuntos de colores disjuntos. Por eso nombramos sus bordes e [i, j] dónde j ∈ [q] es el color del borde mediante una coloración de borde fuerte arbitraria (una coloración en la que no hay dos bordes del mismo color que compartan un vértice) y I ∈ [t] es una indexación arbitraria de la t bordes de cada color. Tenga en cuenta que el t los bordes de cualquier color cubren todos los vértices de D[t, q].

Una prueba del lema anterior aparece en la Sección 5. Tenga en cuenta que D[t, 2] es el gráfico dual de un dispersor estándar.

3. Prueba del factor de inaproximación asintótico por k-SP

Esta sección proporciona una reducción de tiempo polinomial de MAX-3-LIN-q a k-SP. Dada una instancia Φ de MAX-3-LIN-q, es decir, un conjunto de ecuaciones módulo entero q, construimos una instancia de k-SP, es decir, un k-hipergráfico uniforme.

Para el hipergráfico que construimos, agregamos hiperbordes correspondientes a las ecuaciones de Φ y asignaciones satisfactorias a ellas. La idea principal del

La reducción consiste en construir el hipergráfico de tal manera que una gran coincidencia corresponda a una asignación satisfactoria consistente para Φ. Para este propósito, el hipergráfico tiene vértices comunes para los bordes que corresponden a asignaciones que son inconsistentes.

En general, la escasez y uniformidad del gráfico construido están fuertemente relacionadas con la calidad del resultado de dureza. Para obtener un gráfico disperso con un tamaño de borde pequeño, mientras se conservan las propiedades de intersección del borde, utilizamos los gráficos de hiper-borde-dispersor de fi nidos en la sección anterior.

3.1. La construcción.Sea $\Phi = \{\phi_1, \ldots, \phi_{norte}\}$ ser una instancia de MAX-3LIN-q sobre el conjunto de variables X, donde cada variable X ∈ X ocurre una constante numero de veces c (x) = O (1) (como en el teorema 2.3). Describamos ahora cómo para construir, en tiempo polinomial, una instancia de k-SP, el hipergráfico $H_{\Phi} =$ (V, E).

Fijemos, para cada variable X ∈ X, un mapeo uno a uno entre todos índices $I_X \in$ [c (x)] y todas las apariciones de X en Φ.

Recordemos el Lema 2.9 que afirma el hiperdispersor de bordes D[t, q] (Definición 2.8) existe. Para cada variableX ∈ X considera el gráfico D[c (x), q]. Cada vértice en V (D[c (x), q]) corresponde a una ocurrencia de X en Φ, y un número en [D], dónde d = Θ (q Iniciar sesión q). Cada uno de los bordes en MI(D[c (x), q]) está en un uno a uno correspondencia con una ocurrencia $I_X \in$ [c (x)] y un valor a ∈ [q] según la fuerte coloración del borde de D[c (x), q], así que denotemos estos bordes por mi 〈x, yo_X, a〉 .

El conjunto de vértices V de H_{Φ} Consiste en una copia de los vértices del grafo dispersor. D[c (x), q] para cada variable X ∈ X. A saber,

$$V, \{v\langle x, yo, j\rangle \mid X \in X, yo \in [c (x)], j \in [D]\}.$$

De ahora en adelante, para cualquier variable X ∈ X, la copia de D[c (x), q] sobre los vértices V_X , {v 〈x, yo, j〉 | I ∈ [c (x)], j ∈ [D]} será denotado D_X.

Definamos ahora el conjunto de aristas mi de H_{Φ}. Los bordes de H_{Φ} estarán compuesto por varios bordes de los hipergráficos D_X. El conjunto mi consta de un borde para cada ecuación ϕ ∈ Φ sobre variables x, y, z y asignación A a x, y, z eso satisface ϕ. Denotamos por $A|_X, A|_y, A|_z \in [q]$ los valores que A asigna a estas variables (observe que hay q^2 asignaciones tan satisfactorias). Denotamos por I_X, I_y, I_z los índices de las ocurrencias de x, y, z respectivamente en ϕ. El borde correspondiente a ϕ y A es una unión de tres bordes de las copias de hiper-graficas D_X, D_y, D_z de las variables en la ecuación ϕ:

$$mi\langle\phi, A\rangle = mi\langle x, yo_X, A|_X\rangle \cup mi\langle y, yo_y, A|_y\rangle \cup mi\langle z, yo_z, A|_z\rangle .$$

Claramente, la cardinalidad de cada borde mi 〈ϕ, A〉 es 3D, ya que es la unión disjunta de tres aristas de cardinalidad D. Tenga en cuenta que cada uno de los tres bordes mi $\langle x, yo_{x}, A\ |_{x}\rangle$, mi $\langle y, yo_{y}, A\ |_{y}\rangle$, mi $\langle z, yo_{z}, A\ |_{z}\rangle$, que componen mi $\langle \phi, A\rangle$, participa en q en los bordes del hipergrafo H_{Φ}.

En total, los bordes de H_{Φ} están

$$E = \{e\ \langle \phi, A\rangle\ \mid \phi \in \Phi, \text{A es una asignación satisfactoria para } \phi\}.$$

Con esto concluye la construcción del k-Instancia de SP H_{Φ}.

Note que para cada constante q, la construcción se puede realizar en tiempo polinomial determinista. Con este fin, cada dispersorD_{X} debe construirse en tiempo polinomial determinista. Como cada variableX ocurre c (x) = O (1) veces (según el teorema 2.3), el tamaño de D_{X} es constante también. De acuerdo a Lema 2.9, sabemos que D_{X} existe. Por lo tanto, podemos enumerar todos los hipergráficos posibles del tamaño requerido y verificar si son realmente hipergráficos. dispersores de borde con los parámetros requeridos.

3.2. Prueba de corrección.A continuación mostramos que el tamaño de un máximo emparejando en H_{Φ} es proporcional al número máximo de ecuaciones de Φ que pueden satisfacerse simultáneamente. Es decir, si existe una cesión que satisface casi todas las ecuaciones de Φ entonces existe una coincidencia que cubre casi todos los vértices de H_{Φ}. Por otro lado, si cada asignación satisface como máximo una pequeña fracción de las ecuaciones de Φ, entonces cada coincidencia de H_{Φ} es pequeño.

Lema 3.1 (Lo completo). Si hay una asignación para Φ que satisface 1-ε de sus ecuaciones, entonces hay una coincidencia en H_{Φ} de tamaño $((1 - \varepsilon) / q^{2)}\ |E\ |$.

Prueba. Dejar A: X → [q] ser una tarea que satisfaga 1-ε de las ecuaciones. el emparejamiento Considerar METRO ⊆ mi que consta de todos los bordes correspondientes a A, es decir

$$M = \{e\ \langle \phi, A\rangle\ \mid \phi \in \text{Sábado}\ (\Phi, A)\}.$$

Como METRO contiene un borde correspondiente a cada ecuación satisfecha, y para cada ecuación hay q^{2} satisfactorias asignaciones, tenemos $|M\ | = ((1 - \varepsilon) / q^{2)}\ |E\ |$. Para ver que estos bordes son de hecho una coincidencia, considere dos bordes cualesquiera de METRO. Si no se relacionan con las mismas variables, entonces no contienen vértices de un dispersor de hiper-borde conjunto. Por otro lado, si se relacionan con un variable conjunta X ∈ X, luego se relacionan con diferentes ocurrencias $I_{X,1}, I_{X,2} \in [c\ (x)]$, pero la misma tarea a ∈ [q] lo. Por tanto, contienen vértices del mismo dispersor de hiper-borde D_{X}, pero de dos bordes distintos del mismo color, por lo tanto, no comparten un vértice.

Lema 3.2 (Solvencia). Si cada asignación a Φ satisface como máximo un 1 /q + ε fracción de sus ecuaciones, entonces cada coincidencia en H_{Φ} es de tamaño O $(q^{-3}\ |E\ |)$.

La idea de prueba es la siguiente: dada una coincidencia METRO, cada borde en él corresponde a una asignación a tres variables. Dada una coincidenciaMETRO, lo usamos para de fi nir un asignación global $A_{comandante}$ a las variables de Φ: a cada variable se le asigna el valor que concuerda con el número máximo de hiper-bordes de METRO. Luego particionamos los bordes de METRO en dos conjuntos: los que están de acuerdo con la asignación global (llamado $METRO_{comandante)}$ y el conjunto de complementos (llamado $METRO_{min)}$. La talla de METRO $_{comandante}$ está acotado, ya que corresponde al conjunto de ecuaciones satisfechas por A_{maj}cual es pequeña). Luego procedemos a acotar el tamaño de$METRO_{min}$ utilizando la propiedad de expansión de los dispersores de hiper-borde.

Prueba. Denotamos por mi_X los bordes de $H_Φ$ correspondiente a ecuaciones que contienen la variable X,

$$mi_{X} = \{mi \langle \phi, A\rangle \mid \phi \in \Phi (x), e \langle \phi, A\rangle \in MI\}.$$

Denotamos por $mi_{x = a}$ el subconjunto de mi_X correspondiente a una cesión de a a X, es decir,

$$mi_{x = a} = \{mi \langle \phi, A\rangle \mid mi \langle \phi, A\rangle \in mi_{X}, A|_{x} = a\}$$

Dejar METRO ser una coincidencia de tamaño máximo en $H_Φ$. Según el emparejamiento METRO definimos la asignación mayoritaria $A_{comandante}$ de la siguiente manera: para cada X ∈ X, la asignación $A_{comandante}$(X) es el valor a ∈ [q] tal que $|mi_{x = a} \cap M|$ se maximiza. Dejar $METRO_{comandante}$ ser el conjunto de aristas en METRO que está de acuerdo con $A_{comandante}$, y $METRO_{min}$ ser todos los otros bordes en METRO:

$$METRO_{maj} = METRO \cap \{mi \langle \phi, A_{comandante}\rangle \mid \phi \in \Phi\}, METRO_{min} = M \setminus M_{comandante}$$

Como el número de ecuaciones satisfechas por $A_{comandante}$ satisface $|Sat (\Phi, A_{maj})| \leq 1 /q + \varepsilon$, y para cada ecuación hay q^2 aristas correspondientes a todas las asignaciones satisfactorias para esta ecuación, tenemos

$$|METRO_{maj}| < \left(\frac{1}{q} + \varepsilon\right) \frac{|E|}{q^2}. \tag{3,3}$$

A continuación, limitamos el tamaño de $METRO_{min}$. La idea es como sigue: descomponemos cada del borde hacia dentro $METRO_{min}$ en los tres bordes constructivos. Al menos uno de esos tres bordes corresponde a una asignación distinta a la asignación mayoritaria. De ahí que basta con limitar el número de "bordes de construcción" que corresponden a las asignaciones minoritarias. Esto se logra utilizando la propiedad de dispersor.

Considere una determinada variable X ∈ X. Luego D_X es un (q, 1 /q^2) -hiperfilo dispersor (recuerde la definición 2.8). Es decir, en cualquier subconjunto de aristas deD_X que es una coincidencia, todos excepto como máximo un 1 /q^2 fracción de los bordes son de un color (que corresponde a una única asignación a la variable X). Claramente, si dos bordes de D_X se cruzan, entonces también lo hacen cualquier par de bordes de $H_Φ$ que contiene estos dos bordes.

Por lo tanto,

$$(3.4)\qquad \sum_{a6 = A_{comandante}(X)} |METRO_{min} \cap mi_{x=a}| \leq \frac{1}{q^2} |MI(D_X)|$$

donde $|MI(D_x)|$ es el numero de los bordes de D_X.

Considere una ventaja mi $\langle x, yo_X, a\rangle$ de D_X, dónde I_X es el índice de X cuando aparece en la ecuación $\phi \in \Phi$. Este borde se utilizó en la construcción deq bordes de H_Φ, es decir, aquellos que corresponden a las asignaciones satisfactorias de ϕ que asignan a a X.

Por lo tanto, cada borde de D_X es un subconjunto de q hiper-bordes en mi_X. Sin embargo, no más de uno de estos q los bordes pueden participar en M (como METRO es una coincidencia). Taponando esta observación en (3.4) obtenemos

$$(3,5)\qquad \sum_{a6 = A_{comandante}(X)} |METRO_{min} \cap mi_{x=a}| \leq \frac{1}{q^3} |E_X|.$$

Resumiendo sobre la variables de Φ rendimientos

$$(3,6)\qquad |METRO_{min}| \leq \sum_{X \in X,\, a6 = A_{comandante}(X)} |METRO_{min} \cap mi_{x=a}| \leq \frac{1}{q^3} \sum_{X \in X} |mi_X| = \frac{3}{q^3} |E|$$

donde la última igualdad se sigue del hecho de que cada ecuación contiene Tres variables. Así, de (3.3) y (3.6),

$$|M| = |M_{min}| + |METRO_{maj}| \leq \left(\frac{4}{q^3} + \varepsilon\right) |E|.$$

Por los Lemas 3.1 y 3.2 mostramos que Gap-k-SP- $[4/q^3+\varepsilon, 1/q^2-\varepsilon]$ es NP-duro. Dado que cada borde es de tamañok = 3d = Θ (q Iniciar sesión q) es NP-difícil de aproximar k-SP dentro de Ω (k /en k).

4. Ampliación de la prueba para k-DM

La prueba de k-DM es similar al k-SP a prueba, pero tenemos un cuidado adicional para asegurarse de que el gráfico H_Φ que construimos tiene la estructura requerida (es decir, que H_Φ no es sólo k-uniforme, pero también k-fuertemente colorante).

La construcción para k-DM tiene en cuenta la localización de las variables en las ecuaciones en las que aparecen. Como hay tres variables por ecuación, hay tres ubicaciones posibles. Usamos la siguiente notación: Φ (x, l) es el subconjunto de Φ (X) dónde X es el la variable en la ecuación (l $\in$ [3]). Podemos suponer que cada variable aparece el mismo número de veces en cada ubicación.

(Φ (X, 1) = Φ (X, 2) = Φ (X, 3)), ya que podemos tomar tres copias de cada ecuación y cambiar la ubicación de las variables.

Similar a k-SP, asociamos un vértice con cada aparición de una variable. Para cada variableX ∈ X, ahora tenemos tres copias de un dispersor de hiperemaduras (en lugar de solo una que teníamos para k-SP): un dispersor diferente para cada ubicación en las ecuaciones. Para cada lugarl ∈ [3], tenemos un hiper-dispersor D[c (x) /3, q] que se denota por $D_{SG.}$ Los vértices de $H_Φ$ son la unión de los vértices de todos estos hiperdispersores correspondientes a todas las variables en el conjunto de ecuaciones y todas las ubicaciones.

Desde c (x) es exactamente el número de apariciones de la variable X en el conjunto de ecuaciones Φ, podemos enumerar los vértices de $H_Φ$ según la variable X ∈ X y ecuación ϕ ∈ Φ corresponden a (y estos dos parámetros determinar también la ubicación de la variable en la ecuación):

$$V = \{v\langle x, ϕ, j\rangle \mid X ∈ X, ϕ ∈ Φ (x), j ∈ [D]\}.$$

La construcción de los bordes de $H_Φ$ es casi idéntica a la del k-SP, la diferencia es la distinción entre los tres dispersores para cada variable. Observe que hay una biyección entre la ocurrencia de una variable en una determinada ecuación y el vértice correspondiente en uno de los tres hipergráficos correspondientes a esta variable. Por lo tanto, no hay ambigüedad en el proceso de construcción del borde, que por lo demás es idéntico al dek-SP.

La notación que usamos para los bordes es idéntica a la k-Construcción SP también: los bordes corresponden a las asignaciones satisfactorias a las ecuaciones, y se componen de tres bordes dispersores cada uno, mi $\langle ϕ, A\rangle$ = mi $\langle x, yo_{X,} A \mid_X\rangle$ ∪ mi $\langle y, yo_{y,} A \mid_y\rangle$ ∪ mi $\langle z, yo_{z,} A \mid_z\rangle$ (de donde se toman estos tres bordes $D_{X,1,}$ $D_{y,2}$ y $D_{z,3}$ respectivamente). El conjunto de todos los bordes se denota

$$E = \{e\langle ϕ, A\rangle \mid ϕ ∈ Φ, \text{A es una asignación satisfactoria para } ϕ\}.$$

Esto concluye la construcción de k-DM. Primero mostramos que la gráfica construida es de hecho unak-Instancia de DM:

Proposición 4.1. $H_Φ$ es 3D-fuertemente colorante.

Prueba. Mostramos cómo particionar V en 3D conjuntos independientes de igual tamaño. Deja que los conjuntos sean$PAG_{yo, yo}$ por I ∈ [D] y l ∈ [3], donde

$$PAG_{l, yo = \{}v\langle x, ϕ, yo\rangle \mid X ∈ X, ϕ ∈ Φ (x, l)\}.$$

$PAG_{yo, yo}$ es claramente una partición de los vértices, ya que cada vértice pertenece exactamente a una parte.

Ahora explicamos por qué cada parte es un conjunto independiente. Dejar$PAG_{yo, yo}$ ser una parte arbitraria, y dejar mi ⟨ϕ, A⟩ ∈ mi ser un borde arbitrario, donde la ecuación ϕ depende en las variables x, y, z. Por construcción, este borde es una unión disjunta de tres bordes hiperdispersores correspondientes a las tres variables x, y, z,

$$mi\ \langle \phi, A\rangle = mi\ \langle x, yo_{x}, A\ |_{x}\rangle \cup mi\ \langle y, yo_{y}, A\ |_{y}\rangle \cup mi\ \langle z, yo_{z}, A\ |_{z}\rangle .$$

$PAG_{yo, yo}$∩mi ⟨ϕ, A⟩ puede contener vértices correspondientes solo a una de las variables x, y, z, ya que contiene variables correspondientes a una sola ubicación (primero, segundo o tercero). Deje que esa variable sea, digamos,X. Dado que el hipergrafo $D_{X,1}$ es D-uniforme y D-fuertemente colorante, el borde mi ⟨x, yo_{X}, A $|_{X}$⟩ (y por lo tanto mi ⟨ϕ, A⟩) contiene exactamente un vértice de cada uno de los D partes. Por lo tanto, el conjunto$PAG_{yo, yo}$ ∩ mi ⟨ϕ, A⟩ contiene exactamente un vértice. Desde |$PAG_{yo, yo}$ ∩ mi ⟨ϕ, A⟩ | =1 por cada borde y cada conjunto $PAG_{yo, yo}$ la gráfica H_{Φ} es 3D-fuertemente colorante.

Procedemos a probar la brecha en el tamaño máximo de coincidencia entre los casos en el que el conjunto de ecuaciones Φ es casi satisfactorio y muy insatisfactorio. Para el caso en el que Φ es casi satisfactorio, el lema de completitud (Lema 3.1) también es válido para la construcción actual. Demostramos el lema de solidez apropiado, que es muy similar al lema 3.2.

Lema 4.2 (Solvencia). Si cada asignación a Φ satisface como máximo un 1 /q + ε fracción de sus ecuaciones, entonces cada coincidencia en H_{Φ} es de tamaño O (q^{-3} |E |).

Prueba. Denotamos por mi_{SG} los bordes de H_{Φ} correspondiente a ecuaciones ϕ que contiene la variable X en el lugar yo

$$mi_{x, I} = \{mi\ \langle \phi, A\rangle\ |\ \phi \in \Phi (x, I), A \in [q^{2}]\}.$$

Denotamos por $mi_{x = a, I}$ el subconjunto de mi_{SG} correspondiente a una cesión de a ∈ [q] a X, es decir,

$$mi_{x = a, I} = \{mi\ \langle \phi, A\rangle\ |\ \phi \in \Phi (x, I), A\ |_{x = a}\}.$$

Dejar METRO ser una coincidencia de tamaño máximo en H_{Φ}. Según el emparejamiento METRO definimos la asignación mayoritaria $A_{comandante}$, teniendo en cuenta las ubicaciones de las variables, de la siguiente manera: para cada X ∈ X, dejar$\hat{I}$(x) ser la ubicación para la cual $|mi_{x,\hat{I}(x)} \cap M|$ se maximiza. La asignación$A_{comandante}$(X) es el valor a ∈ [q] semejante que $|mi_{x = a,\hat{I}(x)} \cap M|$ se maximiza. Como antes, deja$METRO_{comandante}$ ser el conjunto de aristas en METRO que está de acuerdo con $A_{comandante}$, y $METRO_{min}$ ser todos los otros bordes en METRO:

$$METRO_{maj} = METRO \cap \{mi\ \langle \phi, A_{comandante}\rangle\ |\phi \in \Phi\}, \qquad METRO_{min} = M \setminus M_{comandante}.$$

Exactamente por las mismas razones que en la prueba de solidez anterior. (Lema 3.2), tenemos el estimaciones análogas a (3 (.3) y) (3.5):

(4,3) $|METRO_{aj}| < \left(\frac{1}{q} + \varepsilon\right)\frac{|E|}{q^2}$,

(4,4) $\forall X \in X \quad \sum_{a6 = A_{comandante}(X)}^{metro} |METRO_{min} \cap mi_{x=a,\hat{\Gamma}(x)}| \leq \frac{1}{q^3} |mi_{x,\hat{\Gamma}(x)}|$.

Continuando las líneas del Lema 3.2, nosotros obtener

$$
\begin{aligned}
|M| &\leq \sum_{SG} |METRO \cap mi_{x,l}| \\
&\leq \sum_{SG} |METRO_{comandante} \cap mi_{SG}| + \sum_{x,l,a6 = A_{comandante}(X)} |METRO_{norte} \cap mi_{x=a,l}| \\
&\leq 3 \sum_{X} |METRO_{comandante} \cap mi_{x,\hat{\Gamma}(x)}| + 3 \sum_{x,a6 = A_{comandante}(X)} |METRO_{min} \cap mi_{x=a,\hat{\Gamma}(x)}| \\
&\qquad\qquad \text{(desde } |mi_{x,\hat{\Gamma}(x)} \cap M| \text{ se maximiza por } \hat{\Gamma}(x)) \\
&\leq 3\,|METRO_{maj}| + 3 \sum_{x,a6 = A_{comandante}(X)} |METRO_{min} \cap mi_{x=a,\hat{\Gamma}(x)}| \\
&< 3\left(\frac{1}{q} + \varepsilon\right)\frac{|E|}{q^2} + \frac{3}{q^3}\sum_{X} |mi_{x,\hat{\Gamma}(x)}| \quad \text{(por (4.3) y (4.4))} \\
&\leq \left(\frac{6}{q^3} + 3\varepsilon\right)|E|.
\end{aligned}
$$

Finalmente, por el lema de completitud de la sección anterior y el Lema 4.2, concluimos que Gap-k-DM- $[6/q^3+3\varepsilon, 1/q^2-\varepsilon]$ es NP-hard, por lo que es NP-hard aproximar k-DM dentro de $\Omega(k/\text{en } k)$.

5. Hiperdispersores

En esta sección, probamos el Lema 2.9. Como se indicó anteriormente, los hiperdispersores son generalizaciones de gráficos de dispersores. En la Sección 5.1, demostramos que los parámetros que se dan a continuación son los mejores (hasta una constante) para un hiperdispersor que uno puede esperar lograr.

Lema 2.9. Para cada $q > 1$ y $t > c(q)$ (dónde $c(q)$ es una constante que depende solo de q) existe un hipergrafo $H = (V, E)$ tal que

- $V = [t] \times [D]$, dónde $d = \Theta(q \text{ en } q)$.

- H es un (q, 1 /q^2)-dispersor de hiper-borde.
- H es D-uniforme, D-fuertemente colorante.
- H es q-regular, q-fuertemente colorante en los bordes.

Denotamos este gráfico por D[t, q].

Prueba. Demostramos que la probabilidad de que una gráfica generada aleatoriamente no sea una D[t, q] gráfico es estrictamente menor que 1, lo que da como resultado la existencia de tales gráficos. Dejar

$$V = [t] \times [D]$$

y definir $V_i = [t] \times \{I\}$. A continuación, construimos aleatoriamente los bordes del hipergráfico, de modo que sea D-uniforme y q-regular. Dejar S_t ser el conjunto de todas las permutaciones sobre t elementos. Para cada $(I_1, I_2) \in [q] \times [D]$ elige una permutación de S_t uniformemente en aleatorio:

$$\Pi_{I_1,I_2} \in_R S_t.$$

Definir

(5,1) $$e[i, j] = \{(\Pi_{j,1}(I), 1), (\Pi_{j,2}(I), 2), \ldots, (\Pi_{j,d}(\text{carné de identidad}))\}$$

y deja

$$E = \{e[i, j] \mid (yo, j) \in [t] \times [q]\},$$

entonces $|E| = tq$. Defina una partición de los bordes de la siguiente manera: $mi_i = \{e[j, i] \mid j \in [t]\}$. Así $|mi_1| = \cdots = |mi_q| = t$ y cada conjunto mi_j de aristas cubre cada vértice exactamente una vez. Por lo tanto, H es q-fuertemente colorante en los bordes. Por otro lado, cada borde contiene exactamente un vértice de cada conjunto de vértices V_I. Por lo tanto H es D-fuertemente colorante.

A continuación mostramos que, con alta probabilidad, H tiene la propiedad de dispersor, es decir, cada coincidencia METRO de H se concentra en una sola parte de los bordes, excepto tal vez $(1/q^2)|E| = t/q$ bordes de METRO. Denotamos por PAG la probabilidad que H lo hace no tienen la propiedad de dispersor.

Definit { ion 5.2. ‖Definir

$$METRO_k = \left\{ METRO \subseteq E \mid |M \cap mi_k| \le \frac{t}{q}, |ME_k| = , \forall yo, \frac{t}{q} M \cap E_k \ge |METRO \cap mi_{yo}| \right\}.$$

Proposition 5.3. If H is not a (q, 1/q^2)-hyper-edge-disperser, then there exists a $k \in [q]$ and a set $M \in M_k$ that is a matching.

Proof. Suppose that H is not a $(q, 1/q^2)$-hyper-edge-disperser. Then there exists a matching $M' \subseteq E$ that is not concentrated on one color of edges: $\forall i,\ |M' \setminus E_i| > (1/q^2)|E| = t/q$. Let $k \in [q]$ be such that $|M' \cap E_k|$ is maximal. As any subset of a matching is a matching, we can remove edges from $M' \setminus E_k$ until we are left with exactly t/q edges. Likewise, we can remove edges of $M' \cap E_k$ until this set contains at most t/q edges. Note that the property $\forall i,\ |M' \cap E_k| \geq |M' \cap E_i|$ cannot be violated by the deletion of those edges. Thus the new set obtained is a matching in $\mathcal{M}_k$.

Having the above proposition, we proceed with the proof considering only sets in $\mathcal{M}_1$. Denote by $\Pr[M]$ the probability (over the random choice of H) that M is a matching. By union bound, symmetry with respect to k, and the above proposition,

$$P \leq \Pr_H[\exists k,\ M \in \mathcal{M}_k,\ M \text{ is a matching}] \leq q \sum_{M \in \mathcal{M}_1} \Pr[M] \leq q|\mathcal{M}_1| \Pr[\hat{M}] \tag{5.4}$$

where $\hat{M} \in \mathcal{M}_1$ is the set which maximizes $\Pr[M]$. The size of $\mathcal{M}_1$ is bounded from above by the number of possibilities to choose at most t/q edges from E_1 and another t/q edges from the rest of the edge color sets. Therefore, using the known inequality $\binom{n}{k} \leq (en/k)^k$ and assuming $t \gg q \gg 2$ we obtain

$$|\mathcal{M}_1| \leq \binom{(q-1)t}{t/q}\binom{t+1}{t/q} \leq (eq^2)^{t/q}(eq)^{2t/q} \leq (eq)^{4t/q}. \tag{5.5}$$

We next bound $\Pr[\hat{M}]$. Denote by $\hat{M}_i$ the event that $\hat{M}$ restricted to the vertices of V_i is a matching (that is, the edges of $\hat{M}$ do not share a vertex in V_i). According to the independent choice of permutations in the construction of H (recall (5.1)), the events $\hat{M}_i$ are independent and identically distributed. Hence,

$$\Pr[\hat{M}] = \prod_{i=1}^{d} \Pr[\hat{M}_i] \tag{5.6}$$

and we proceed to bound $\Pr[\hat{M}_1]$. Henceforth we shall only consider vertices of V_1.

Let M_i be the set of edges in $\hat{M} \cap E_i$ restricted to the vertices of V_1. Let A_i be the event that the sets of edges $\{M_j \mid j \leq i\}$ are all disjoint. Then

$$\Pr[\hat{M}_1] = \Pr\Big[\bigcap_{i=2}^{q} A_i\Big] = \prod_{i=2}^{q} \Pr[A_i \mid A_{i-1}]. \tag{5.7}$$

The probability of the event $A_i \mid A_{i-1}$ is the probability of picking at random $|M_i|$ different vertices from a set of t vertices (the set V_1), and avoiding all vertices from $\bigcup_{l=1}^{i-1} M_l$. Naturally, this probability is smaller than the probability of picking $|M_i|$ vertices from a set of t vertices with repetition (one is allowed to choose the same vertex more than once). The assumption A_{i-1} implies that the sets M_l for all $l < i$ are disjoint, and hence $|\bigcup_{l=1}^{i-1} M_l| = \sum_{l=1}^{i-1} |M_l|$. Therefore,

$$\Pr[A_i \mid A_{i-1}] \le \left(1 - \frac{\sum_{l<i} |M_l|}{t}\right)^{|M_i|} \le e^{-\frac{1}{t}|M_i| \sum_{l<i} |M_l|}$$

where for the last inequality we used $1 - x \le e^{-x}$. Thus by (5.6) and (5.7) we have

$$\Pr[\hat{M}] \le e^{-(d/t) \sum_{i=2}^{q} \left(|M_i| \sum_{j=1}^{i-1} |M_j|\right)} = e^{-(d/t) \sum_{i<j} |M_i||M_j|}. \tag{5.8}$$

We need an upper bound on the previous probability, and that is obtained when the term $\sum_{i<j} |M_i||M_j|$ is minimized. In our case, the constraint that $\hat{M} \in \mathcal{M}_1$ implies that $|M_1| \ge \max_{i=2}^{q} |M_i|$ and $\sum_{i=2}^{q} |M_i| = t/q$. Lemma 5.10 below shows that the minimum of this expression under these constraints is at least $t^2/4q^2$. Therefore, from 5.8, we obtain the following bound on the probability:

$$\Pr[\hat{M}] \le e^{-\frac{d}{t} \cdot \frac{t^2}{4q^2}} = e^{-\frac{dt}{4q^2}}. \tag{5.9}$$

Therefore by (5.4), (5.5), (5.9),

$$P \le q(eq)^{4t/q} e^{-dt/4q^2}$$

Any d which guarantees that $q(eq)^{4t/q} e^{-dt/4q^2} \ll 1$ suffices to conclude that $P < 1$, and therefore that there exists H with the disperser properties. Simple calculations show that we require $d > (4q^2 \ln q)/t + 12q(1 + \ln q)$. Since $t > q \ge 2$, any $d \ge 100q \ln q$ suffices.

It remains to prove the following technical lemma:

Lemma 5.10. Under the constraints

$$\forall i \in [m],\ x_i \ge 0, \qquad x_1 \ge \max_{i=2}^{q} x_i, \qquad \sum_{i=2}^{q} x_i = T,$$

we have

$$\sum_{1 \le i < j \le q} x_i x_j \ge \frac{1}{4} T^2.$$

Proof. If $x_1 \ge T/2$, then we directly obtain

$$\sum_{1\le i<j\le q} x_i x_j \ge x_1 \sum_{i=2}^{q} x_i \ge \frac{T}{2}\cdot T \ge T^2\frac{1}{4}.$$

Otherwise, we know that

$$\sum_{1\le i<j\le q} x_i x_j \ge \sum_{2\le i<j\le q} x_i x_j = \frac{1}{2}\Big(\sum_{i=2}^{q} x_i\Big)^2 - \frac{1}{2}\sum_{i=2}^{q} x_i^2 \ge \frac{1}{2}T^2 - \sum_{i=2}^{q} x_i^2.$$

The function $\sum_{i=2}^{q} x_i^2$ is convex, and hence under the constraints $\sum_{i=2}^{q} x_i = T$ and $\max_{i\ge 2} x_i \le T/2$, it is maximized where $x_2 = x_3 = T/2$ and the rest of the variables are zero. We obtain $\sum_{i=2}^{q} x_i^2 \le \frac{1}{4}T^2$, and finally

$$\sum_{1\le i<j\le q} x_i x_j \ge T^2\frac{1}{2} - \sum_{i=2}^{q} x_i^2 \ge \frac{T^2}{4}.$$

5.1. Optimality of hyper-disperser construction. We now turn to see why the hyper-disperser from Lemma 2.9 has optimal parameters. We base our observation on the lemma below from Radhakrishnan & Ta-Shma (2000):

Definition 5.11. A bipartite graph $G = (V_1, V_2, E)$ is called a δ-disperser if for every $U_1 \subseteq V_1$ and $U_2 \subseteq V_2$ with $|U_1|, |U_2| \ge \delta|V_1| = \delta|V_2|$, the subset $U_1 \cup U_2$ is not an independent set.

Lemma 5.12 (Radhakrishnan & Ta-Shma 2000). *Every bipartite d-regular 1/k-disperser must satisfy $d = \Omega(k \ln k)$.*

Using this lemma we prove:

Lemma 5.13. *Every d-uniform, d-strongly-colorable, q-regular, q-stronglyedge-colorable $(q, 1/q^2)$-hyper-edge-disperser must satisfy $d = \Omega(q \ln q)$.*

Proof. Consider a hyper-disperser $H = (V_H, E_1, \ldots, E_q)$ as in the statement. Let us construct a bipartite graph $G = (V_1, V_2, E_G)$ as follows. Let

$$V_1 = E_1, \qquad V_2 = E_2,$$
$$E_G = \{(e_i, e_j) \mid e_i \in E_1, e_j \in E_2, e_i \cap e_j \neq \emptyset\}.$$

The graph G is bipartite since the edge sets E_1 and E_2 of H are nonintersecting (as H is q-strongly-edge-colorable). To conclude that G is also d-regular observe the following: H is d-uniform, therefore, every edge $e \in E_1 \cup E_2$

contains d vertices. Moreover, H is q-regular, q-strongly-edge-colorable, thus every vertex of H is contained once in an edge of E_1 and once in an edge of E_2. Therefore, every edge of E_1 intersects one edge of E_2 for each of its d vertices (and vice versa). Thus, every vertex of G is of degree d.

In addition, G is a disperser: consider any two sets $S_1 \subseteq V_1$ and $S_2 \subseteq V_2$ of size $|S_1| = (1/q)|V_1|$ and $|S_2| = (1/q)|V_2|$. The corresponding sets (of edges) in H are of fractional size $1/q^2$ each, thus, as H is a $(q, 1/q^2)$-hyper-edge-disperser, they contain intersecting edges implying that $S_1 \cup S_2$ is not an independent set in G. Since this is true for any $S_1 \subseteq V_1$ and $S_2 \subseteq V_2$, G is a 1/q-disperser.

As G is a bipartite d-regular 1/q-disperser, $d = \Omega(q \ln q)$ by Lemma 5.12.

6. Discussion

An interesting property of our construction is the almost perfect completeness. This property refers to the fact that the matching proved to exist in the completeness lemma 3.1 is almost perfect, that is, it covers $1 - \varepsilon$ of the vertices. Knowing the location of a gap is interesting in its own right and may prove useful (in particular if it is extreme on either the completeness or the soundness parameters, see for example Petrank 1994). In fact, applying our reduction on other PCP variants instead of Max-3-Lin-q (e.g. parallel repetition of 3-SAT) yields perfect completeness for k-SP and for k-DM (but with weaker hardness factors).

The ratio between the asymptotic inapproximability factor presented herein for k-SP and k-DM, and the tightest approximation algorithm known was reduced to $O(\ln k)$. The open question of where in the range, from k/2 to $O(k/\ln k)$ is the approximability threshold, is of independent interest, as its implications to the difference between k-DM and k-IS. The current asymptotic inapproximability factor of $\Omega(k/\ln k)$ for k-DM approaches the tightest approximation ratio known for k-IS, namely $O(k \log\log k/\log k)$ by Vishwanathan (1996). Thus, a small improvement in either the approximation ratio or the inapproximability factor will show these problems to be of inherently different complexity.

Acknowledgements

We would like to thank Adi Akavia and Dana Moshkovitz for their helpful comments.

References

N. Alon, U. Fiege, A. Wigderson & D. Zuckerman (1995). Derandomized graph products. Comput. Complexity 5, 60–75.

S. Arora, C. Lund, R. Motwani, M. Sudan & M. Szegedy (1998). Proof verification and intractability of approximation problems. J. ACM 45, 501–555.

S. Arora & S. Safra (1998). Probabilistic checking of proofs: a new characterization of NP. J. ACM 45, 70–122.

R. Bar-Yehuda & S. Moran (1984). On approximation problems related to the independent set and vertex cover problems. Discrete Appl. Math. 9, 1–10.

P. Berman & T. Fujito (1995). On approximation properties of the independent set problem for degree 3 graphs. In Algorithms and Data Structures (Kingston, ON), Lecture Notes in Comput. Sci. 955, Springer, 449–460.

P. Berman & M. Furer (1994). Approximating maximum independent set in bounded degree graphs. In Proc. 5th Annual ACM-SIAM Symposium on Discrete Algorithms (Arlington, VA), 365–371.

P. Berman & M. Karpinski (2003). Improved approximation lower bounds on small occurrence optimization. Electronic Colloq. on Comput. Complexity (ECCC) 10 (008).

R. Boppana & M. M. Halldórsson (1992). Approximating maximum independent sets by excluding subgraphs. BIT 32, 180–196.

J. Edmonds (1965). Paths, trees and flowers. Canad. J. Math. 17, 449–467.

U. Feige, S. Goldwasser, L. Lovász, S. Safra & M. Szegedy (1996). Interactive proofs and the hardness of approximating cliques. J. ACM 43, 268–292.

M. M. Halldórsson (1998). Approximations of independent sets in graphs. In Approximation Algorithms for Combinatorial Optimization (Aalborg), Lecture Notes in Comput. Sci. 1444, Springer, 1–13.

J. H˚ åstad (1999). Clique is hard to approximate within $n^{1-\varrho}$. Acta Math. 182, 105–142.

J. H˚ åstad (2001). Some optimal inapproximability results. J. ACM 48, 798–859.

E. Hazan (2002). On the hardness of approximating k-dimensional matching. Master's thesis, Tel-Aviv Univ.

C. A. J. Hurkens & A. Schrijver (1989). On the size of systems of sets every t of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems. SIAM J. Discrete Math. 2, 68–72.

V. Kann (1991). Maximum bounded 3-dimensional matching is MAXSNP-complete. Inform. Process. Lett. 37, 27–35.

R. M. Karp (1972). Reducibility among combinatorial problems. In Complexity of Computer Computations (Yorktown Heights, NY), Plenum, 85–103.

M. Mucha & P. Sankowski (2004). Maximum matchings via gaussian elimination. In Proc. 45nd IEEE Symp. on Foundations of Computer Science, 248–255.

C. Papadimitriou (1994). Computational Complexity. Addison, Wesley.

E. Petrank (1994). The hardness of approximation: gap location. Comput. Complexity 4, 133–157.

J. Radhakrishnan & A. Ta-Shma (2000). Bounds for dispersers, extractors, and depth-two superconcentrators. SIAM J. Discrete Math. 13, 2–24.

R. Raz (1998). A parallel repetition theorem. SIAM J. Comput. 27, 763–803.

L. Trevisan (2001). Non-approximability results for optimization problems on bounded degree instances. In Proc. 33rd ACM Symp. on Theory of Computing, 453–461.

S. Vishwanathan (1996). Personal communication to M. Halld´órsson cited in Halldórsson (1998).

A. Wigderson (1983). Improving the performance guarantee for approximate graph coloring. J. ACM 30, 729–735.

Manuscript received 24 December 2003

Elad Hazan
Computer Science Department
Princeton University
Princeton, NJ 08540, U.S.A.
www.cs.princeton.edu/~ehazan

Shmuel Safra
School of Computer Science
Tel Aviv University
Tel Aviv 69978, Israel
safra@post.tau.ac.il

Oded Schwartz
School of Computer Science
Tel Aviv University
Tel Aviv 69978, Israel
odedsc@post.tau.ac.il